DOCUMENT RESUME

ED 238 941 TM 840 043

TITLE	The Role of New Technologies in Schools: Collected Papers. Research Into Practice Project.
INSTITUTION	California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY	National Inst. of Education (ED), Washington, DC.
PUB DATE	Nov 83
GRANT	NIE-G-83-0001
NOTE	155p.; Papers presented at the Paths to Excellence: Testing and Technology Conference (Los Angeles, CA, July 14-15, 1983). For related document, see TM 840 051.
PUB TYPE	Collected Works - Conference Proceedings (021) -- Information Analyses (070)
EDRS PRICE	MF01/PC07 Plus Postage.
DESCRIPTORS	Cognitive Processes; Computer Assisted Instruction; *Computer Assisted Testing; Computer Programs; *Educational Policy; *Educational Technology; *Educational Trends; *Futures (of Society); *Instructional Design; Instructional Development; Media Selection; Microcomputers; School Districts; State Boards of Education; Teacher Effectiveness; Test Construction; Testing

ABSTRACT

The Center for the Study of Evaluation of the Graduate School of Education at the University of California at Los Angeles hosted a two-day conference on "Paths to Excellence: Testing and Technology" on July 14-15, 1983. Attended by over 100 educational researchers, practitioners, and policymakers, day one of the conference focused on issues in educational testing; day two explored the status and future of technology in schools. This document presents the collected papers from the second day of the conference. Presenters representing a broad range of disciplines and local, state, and national policy perspective were asked to consider issues in technology in the schools and the policy implications of present and future applications. Presenters were given broad topic areas: for example, human cognition, instructional design, test design, software evaluation, and social policy. Their charge was to explore their topic areas in light of new technologies with regard to the following: (1) What is the current state of the art? (2) What are potential future directions? (3) What barriers may impede future directions? and (4) What are the implications for educational research, policy, and practice? (PN)

RESEARCH INTO PRACTICE PROJECT

THE ROLE OF NEW TECHNOLOGIES IN SCHOOLS: COLLECTED PAPERS

Eva L. Baker and Joan L. Herman

Project Directors

Grant Number:

NIE-G-83-0001

Center for the Study of Evaluation

UCLA Graduate School of Education

Los Angeles, California

November 1983

The project presented or reported herein was performed pursuant to a grant from the National Institute of Education, Department of Education. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

Contents

Introduction

The UCLA Center for the Study of Evaluation (CSE) hosted a two-day conference on "Paths to Excellence: Testing and Technology" on July 14-15, 1983. Attended by over 100 educational researchers, practitioners, and policymakers, day one of the conference focused on issues in educational testing; day two explored the status and future of technology in schools.

This document presents the collected papers from the second day of the conference. Presenters representing a broad range of disciplines and local, state, and national policy perspective were asked to consider issues in technology in the schools and the policy implications of present and future applications. Presenters were given broad topic areas: for example, human cognition, instructional design, test design, software evaluation, and social policy. Their charge was to explore their topic areas in light of new technologies with regard to the following:

1. What is the current state of the art?
2. What are potential future directions?
3. What barriers may impede future dirctions?
4. What are the implications for educational research, policy, and practice?

Human Cognition and the Use of New Technologies

Richard E. Mayer

Psychology Department

University of California, Santa Barbara

Introduction

Objective. Computer technology is invading our nation's schools. However, the ultimate usefulness of this new technology may be viewed with either optimisim or pessimism. In the optimistic view, computers will becomes aides for teachers, providing help in areas such as instruction, problem solving, and evaluation. In the pessimistic view, computers will become an expensive fad and eventually join their predecessors--teaching machines--collecting dust in the basements of schoolhouses across the nation.

The purpose of this talk is to convince you that the effective use of computer technology in schools requires an understanding of how humans learn and think. The fulfillment of the optimistic scenario of computers depends on their being used in a way that is consistent with what we know about the psychology of human cognitive processes. In order to avoid the pitfalls of the past, and thus to deny the fulfillment of the pessimistic scenario, we must not base the use of computer technology on psychological principles which are inappropriate.

Rationale. The tremendous influx of computer technology into our nation's schools has been widely reported. In a recent report to school board members, Fortune (1983) points out that more than 100,000 microcomputers and terminals were installed in schools in 1982, and that there will be almost one million microcomputers in schools by

1985. Similarly, a recent report in News (1983) stated:

> As of last spring, by one count, 29,000 schools provided... microcomputers and terminals for 4,711,000 school students. Another study released last fall found that 60 per cent of the nation's school districts use computers for learning and that the number of elementary schools using them had increased 80 percent over the year before. In fact, computers are multiplying too fast to count; experts figure the statistics are obsolete when they are reported.

In California, the Apple Computer Foundation's "Kids Can't Wait" program is providing one computer system for every school in the state, and the state's "Investment in People" program is providing about $10,000,000 for the improvement of education related to "high technology". Fortune (1983, p. 7) summarizes all of the new programs as follows: "One thing is clear: computers in the school are not just a passing fad."

The urgent need to prepare for the role of computers in schools has been widely recognized. For example, a recommendation from Technology in Science Education: The Next Ten Years (National Science Foundation, 1979) states that "there is an urgent, national need to create an educational system that fosters computer literacy in our society." The report points out that "American education is not only missing a great opportunity, it is failing to discharge a crucial responsibility" (Deringer & Molnar, 1982).

As another example, the President's Report on Science and Engineering Education in the 1980's and Beyond (National Science

Foundation, 1980) cites the decline in national productivity and increase in foreign trade competition as rationale for preparing American students to become better educated in the use of computers. The French government has recognized the impending "computerization of society" and has committed France to a national policy of computer education for all students (Nora & Minc, 1980). In addition, state departments of education in this country have begun to propose computer courses as part of the mandated graduation requirements (California State Department of Education, 1982).

A recent conference on National Goals for Computer Literacy in 1985 (Seidel, Anderson & Hunter, 1982) concluded by calling for "the presence of computers for instruction in all schools for all students" and "the availability of a critical mass of high-quality curricula and courseware." In particular, the conference supported the proposition that a computer should be in every classroom from kindergarten through eighth grade; in grades 8 through 12, computers should be available in a laboratory environment for every student."

The National Council of Teachers of Mathematics (1980) has issued similar recommendations in its report An Agenda for Action: Recommendations for Mathematics of the 1980's. One recommendation concerning computers stated: "Mathematics programs should take full advantage of the power of calculators and computers at all grade levels." More specifically, the report states, "All high school students should have work in computer literacy and hands-on use of computers."

Two Scenarios

The foregoing section demonstrated that computer technology is

arriving in our schools. Let me try to describe two scenarios for the role of computers in improving our children's education: a pessimistic scenario and an optimistic scenario.

In order for you to fully appreciate the pessimistic scenario for the future, I ask that you consider the past history of technology in the schools. In particular, let's briefly review the role of teaching machines in education, and the theory of learning and instruction which supported their use.

Teaching machines clattered onto the scene of American education about 25 years ago (Skinner, 1958). In his classic book The Technology of Teaching Skinner (1968, p.22) introduced an early version of a teaching machine:

> The device is a box about the size of a small record player. On the top surface is a window through which a question or problem printed on paper tape may be seen. The child answers the question by moving one or more sliders upon which the digits 0 through 9 are printed. The answer appears in square holes punched in the paper upon which the question is printed. When the answer has been set, the child turns a knob. The operation is as simple as adjusting a television set. If the answer is right, the knob turns freely and can be made to ring a bell...If the answer is wrong, the knob will not turn. When the answer is right, a further turn of the knob engages a clutch which moves the next problem into place in the window.

Some more sophisticated versions of teaching machines involved answer keys instead of knobs, and even allowed the students to write an answer.

From the beginning, the technological development of teaching machines was closely tied to an underlying theory of human learning. The dominant force in psychology at the time was behaviorism. Hence, the principles of learning by reinforcement guided the use of teaching machines. In particular, the primary instructional materials for teaching machines were teaching programs--a series of simple questions, each requiring an overt response from the learner. For example, a program in high school physics began with the following items (Skinner, 1968, p. 45):

> The important parts of a flashlight are the battery and the bulb. When we "turn on" a flashlight, we close a switch which connects the battery with the _____.

> When we turn on a flashlight, an electric current flows through the fine wire in the ____ and causes it to grow hot.

> When the hot wire glows brightly, we say that it gives off or sends out heat and _____.

For each item, the student fills in the missing word, and then uncovers the corresponding word or phrase. In the above example, the correct answers respectively are: bulb, bulb, and light. As you can see, the instructional materials are based on the idea that learners must make a response, and that the response must be immediately reinforced.

Skinner's arguments for bringing teaching machines into schools are remarkably similar to many current arguments for using computers

in schools. For example, Skinner (1968, p.26) notes that new technology will aid rather than replace the teacher: "The changes proposed should free her for the effective exercise of her (teaching)." Similarly, Skinner (1968, p. 27) addresses the issue of cost: "Can we afford to mechanize our schools? The answer is clearly Yes."

In spite of the early enthusiasm of Skinner and many others, teaching machines did not revolutionize education. This failure to "mechanize teaching" motivates the questions: Will the computers being introduced today soon join their teaching machine predecessors, collecting dust in schoolhouse basements? Will computers, like teaching machines, fail to live up to the claims that have been made for them, and instead become just another costly fad in education? Twenty-five years from now, will we look back on Papert's (1980, p. 13) observation that "very powerful kinds of learning" take place with computers in the same way we now smile at Skinner's (1968, p. 28) claim that "the equipment needed (for educational innovation) can easily be provided"?

Proponents of the pessimistic scenario may answer "yes" to these questions. In the pessimistic scenario, computers do not find a home in American schools. Yet, there are several factors which lessen the appeal of the pessimistic scenario. First, the computer technology of today is far more powerful than the teaching machine technology of 25 years ago. Computers are not constrained by having to provide a series of test items; instead; computers allow for storage of massive data bases, graphics and simulations, interactive communication, and so on. Second, the current state of psychology has changed

dramatically over the past 25 years. The behaviorist theories of learning, based largely on animal research, have been replaced by cognitive psychology. Cognitive psychology provides implications for the instructional use of computer technology that are very different from earlier behaviorist-inspired instructional materials.

In the optimistic scenario, modern theories of learning and cognition are used in the development of useful instructional materials for computers. For example, cognitive psychologists tend to view learning as the acquisition of knowledge rather than the acquisition of responses. Mayer (1981) has shown how the analytic theories of cognitive psychology have been applied to several kinds of knowledge:

semantic knowledge--factual knowledge about the world, such as rainfall patterns for South America.

procedural knowledge--knowledge about how to carry out some procedure, such as how to compute in long division.

strategic knowledge--knowledge about how to set goals and monitor progress towards solving a problem, such as how to plan the writing of a research paper.

One of the major accomplishments of cognitive psychology has been the development of techniques for precisely describing each of these kinds of knowledge within specific domains (Mayer, 1981). These techniques have implications for how to design effective instructional uses of computers. In the remainder of this paper, examples are given of possible uses of computers to enhance acquisition of each type of knowledge.

The Computer as an Aid to Learning Semantic Knowledge

Semantic knowledge refers to a person's factual knowledge about the world. Examples include knowledge about geography, such as how climate and terrain are related to a region's major crops, or the determinants of the amount of rainfall in a region.

Recent research on the psychology of human learning and cognition suggests a different approach to instruction as compared to the behaviorist approach which dominated during the teaching machine revolution. These differences can be summarized as follows:

active understanding versus passive memorization--The cognitive approach views learning as an active process in which the learner searches for meaning in what is presented, rather than a passive process of performing and remembering what the instructor demands.

assimilative versus additive--The cognitive approach views learning as a process of connecting new information with existing knowledge structures, rather than adding isolated pieces of information to memory.

cognitive structures versus responses--The cognitive approach views the outcome of learning as a coherent body of knowledge (or "mental model") rather than a set of specific responses for specific stimuli.

If meaningful learning of semantic knowledge is an active process of assimilating and reorganizing information, then computers may be used in a way that encourages active exploration. For example, Collins & Stevens (1982) have developed an "intelligent tutor" that uses an inquiry or Socratic method, and that can be used with existing computers. The system is based on the idea that learning about some

new domain, such as geography or meteorology, involves the construction of a "mental model" which relates all of the variables in the system.

Based on the observations of good human tutors, Collins (1977) developed rules for how to engage in inquiry teaching. Some of the main rules for how to teach are summarized below:

1. Ask about a known case, such as "Do they grow rice in China?"
2. Ask for any factors, such as "Why can they grow rice in China?"
3. Ask for intermediate factors, such as "Why do monsoons make it possible to grow rice in China?"
4. Ask for prior factors, such as "What do you need to have enough water?"
5. Form a general rule for an insufficient factor, such as "Do you think any place with enough water can grow rice?"
6. Pick a counterexample for an insufficient factor, such as "Why don't they grow rice in Ireland?"
7. Form a general rule fo. an unnecessary factor, such as "Do you think it is necessary to have heavy rainfall in order to grow rice?"
8. Pick a counterexample for an unnecessary factor, such as "Why do they grow rice in Egypt when they don't have much rainfall?"

Collins and Stevens (1982) have summarized the strategies that an intelligent tutor should use in teaching a student. Some strategies involve selecting a case, and then using counterexamples. An example

of this strategy is demonstrated in the following dialogue (Collins & Stevens, 1982, p. 81):

Tutor: Why do they grow rice in Louisiana?

Student: Places where there is a lot of water. I think rice requires the ability to selectively flood fields.

Tutor: O.K. Do you think there's a lot of rice in, say, Washington and Oregon?

Collins's and Stevens's tutor requires a lot of specific knowledge (such as knowledge about geography), as well as procedures for asking questions and strategies for organizing the questions.

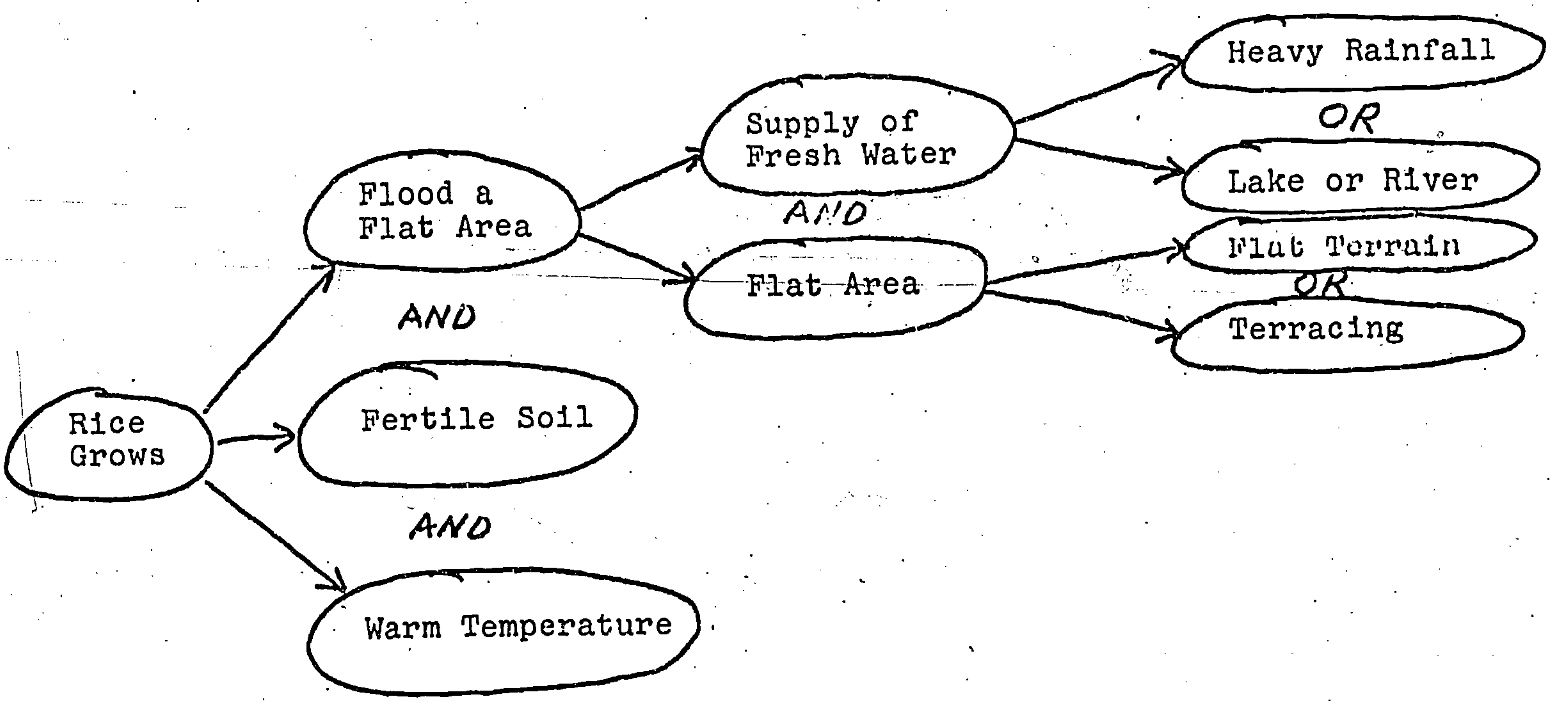
What is learned from a computerized tutor such as the one proposed by Collins and Stevens? A student may form a mental model of the factors involved in growing rice, such as summarized in Figure 1. As you can see, the student builds a coherent structure of factors and relations rather than a set of specific factual answers to specific questions. The mental model allows the student to generate answers to novel questions, and may be used in learning new information.

The use of computers as Socratic tutors represents an exciting possibility, especially in situations where the goal is to teach semantic knowledge. However, the main point in my example is that the way in which the computer is used is determined by the underlying theory of human learning and cognition that is currently available. Thus, the success or failure of computer technology in teaching semantic knowledge depends as much on the educational implications of cognitive psychology as on the power of computer technology itself.

The Computer as an Aid to Learning Procedural Knowledge

Procedural knowledge refers to a person's knowledge about how to

Figure 1. Factors Influencing the Growing of Rice



do something. Examples include knowledge about how to carry out long division or three-digit subtraction. The cognitive approach to procedural knowledge is based on analyzing any procedure into its parts. According to the cognitive approach, the description of procedural knowledge is based on what is learned rather than on how much is learned. Instead of focusing on the percentage of correct answers, the cognitive approach focuses on describing the procedure that the student is using to generate the answers.

Cognitive psychologists have been successful in analyzing many mathematical tasks into their constituent parts. For example, Groen and Parkman (1972) have described several different procedures that children might use to solve problems of the form m + n (where the sum is less than 10). The models are based on the idea that the child uses counting as a way of finding answers to addition problems. Three possible procedures are:

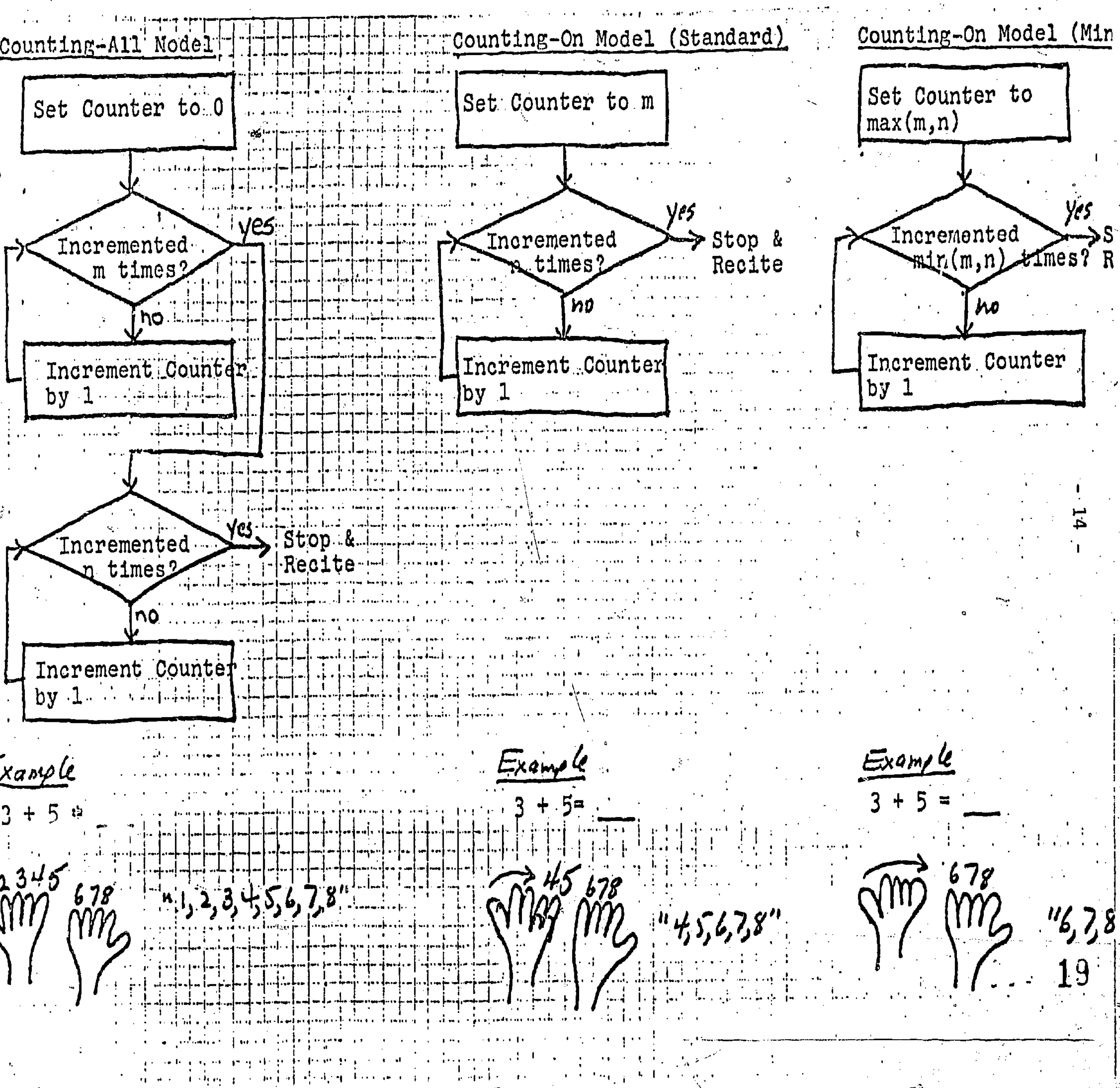
counting-all--Set a counter to 0. Increment it m times and then increment it n times. For 3 + 5, the child recites, "1,2,3...4,5,6,7,8."

counting-on--Set a counter to the first number (m); increment it n times. For 3 + 5, the child states, "4,5,6,7,8."

min model (for counting-on)--Set a counter to the larger of m or n; increment the counter by the smaller of m or n. For 3 + 5, the child states, "6,7,8."

Examples of these three procedures are given in Figure 2; the diamonds represent decisions and the rectangles represent operations. Fuson (1982) has observed a developmental progression in which children move from a counting-all procedure to a counting-on

Figure 2. Three Counting Models of Simple Addition



procedure, and eventually to a known-facts procedure in which the answers are memorized.

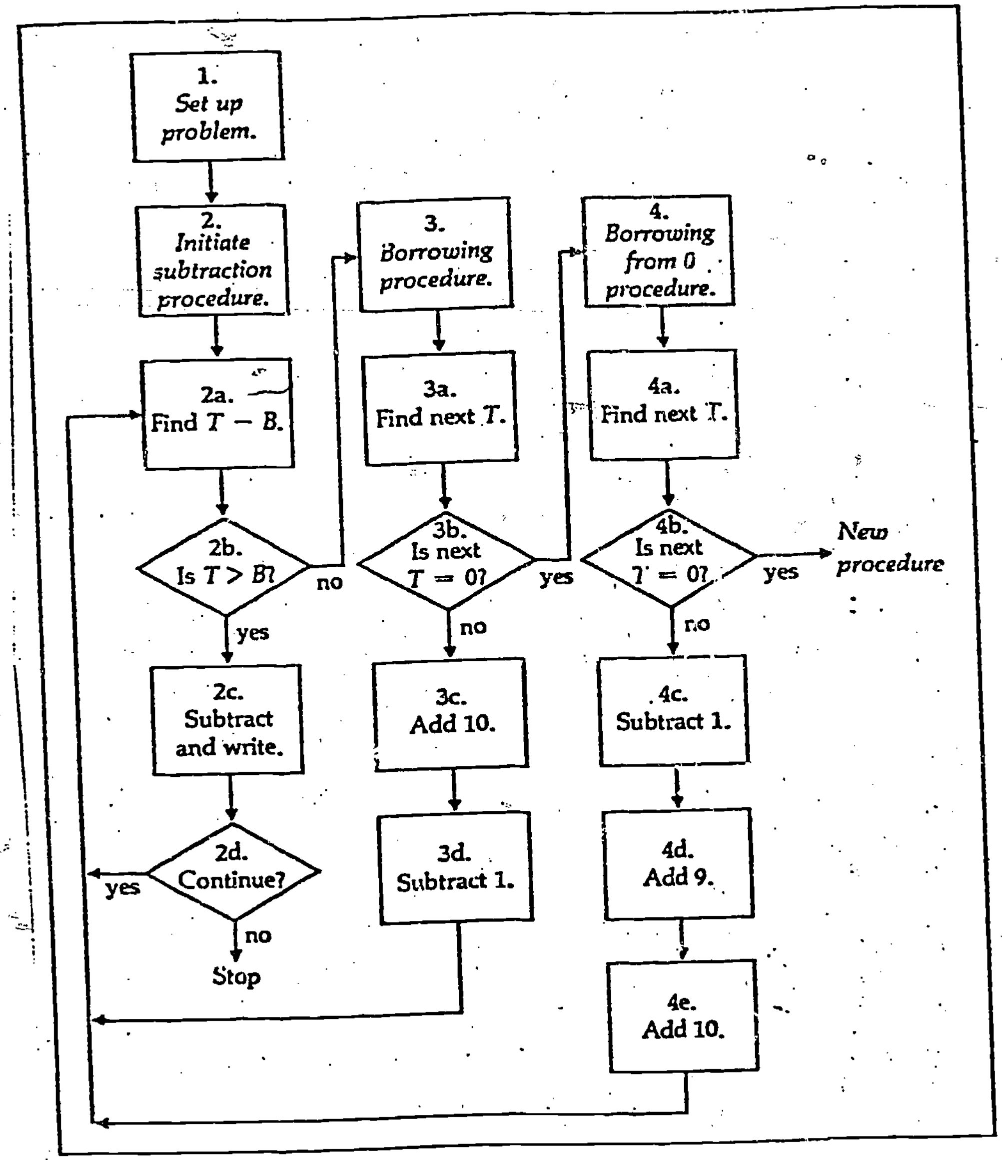
A slightly more complex computational task is three-digit subtraction, such as 697 - 354 =...... Figure 3 shows a computational procedure which generates correct answers for three-digit subtraction problems. If a student possesses this knowledge, then the student will be able to generate correct answers for all three-digit subtraction problems. However, suppose that a student gives answers such as below:

521	819	712	481	655
-418	-203	-531	-380	-160
117	616	221	101	515

We could describe this student's performance by saying that he is right on 40% of the problems. However, a more useful approach is to try to describe the procedure that the student is following. For example, we could say that this student is using the procedure in Figure 3, but with small "bugs"; namely, at steps 2a, 2b, and 2c, the student subtracts the smaller number from the larger number regardless of which is on top.

Brown and Burton (1978) have argued that students' computational performance can be described by saying that they are using a procedure--perhaps with some bugs in it--and applying this procedure consistently to problems. In order to test this idea, Brown and Burton (1978) gave a set of 15 subtraction problems to 1,325 primary school children. Brown and Burton developed a computer program called BUGGY to analyze each student's procedural algorithm for three-digit subtraction. If the student's answers were all correct, BUGGY would

Figure 3. A Process Model for Three Column Subtraction



categorize that subject as using the correct algorithm. If there were errors, BUGGY would attempt to find one bug that could acount for all or most of the errors. If no single bug could account for the errors, then all possible combinations were tried, until BUGGY found combinations that best accounted for the errors. Figure 4 lists some of the most common bugs, such as "borrowing from zero" or subtracting smaller from larger". The BUGGY program was able to describe the performance of about half of the students by providing a list of each student's "bugs". Thus, Brown's and Burton's work provides a means for pinpointing specific bugs in students' computational procedures.

The BUGGY program provides an example of how computer technology can be used to improve the teaching of procedural knowledge. The BUGGY program provides the teacher with a detailed diagnosis of errors in "what is learned" so that the student can be given instruction aimed specifically at remediating the bugs. Again, my point is that the use of computers in teaching of procedural knowledge can be closely guided by existing theories in cognitive psychology.

The Computer as an Aid to Learning Strategic Knowledge

Strategic knowledge refers to knowledge concerning how to set goals, select procedures for achieving goals, and monitor progress toward goals. Examples include knowledge of how to plan the writing of a research paper or how to produce a computer program that accomplishes some task. Research in cognitive psychology emphasizes the role of <u>process</u> rather than <u>product</u> in creative problem solving. For example, consider the following assignments: "Write an essay on whether children should be allowed to choose their own courses in school" or "Write a BASIC program that will take a list of names as

Figure 4. Some Common Subtraction Bugs

Number of Occurrences in 1325 Students	Name	Example	Description
57	Borrow from zero	103 − 45 = 158	When borrowing from a column whose top digit is 0, the student writes 9, but does not continue borrowing from the column to the left of zero.
54	Smaller from larger	253 − 118	The student subtracts the smaller digit in each column from the larger, regardless of which one is on top.
10	Diff 0−N=N	140 − 21	Whenever the top digit in a column is 0, the student writes the bottom digit as the answer.
34	Diff 0−N=N and move over zero	304 − 75	Whenever the top digit in a column is 0, the student writes the bottom digit as the answer. When the student needs to borrow from a column whose top digit is zero, he skips that column and borrows from the next one.

input and give an alphabetized list as output." Instruction could focus on the final product, such as a holistic rating of the final essay or whether the BASIC program runs properly, or could focus on the processes that a person went through in generating the final product, including setting of goals, etc.

Research on the process of writing (Hayes & Flower, 1980) has identified the following processes in writing: planning, in which the author searches memory for ideas and uses these ideas to establish a plan for generating text; translating, the actual production of text; and reviewing, the improvement of the written text. According to these researchers, writing may be viewed as a problem-solving process in which goals are set and monitored.

How can the computer become involved as an aid in writing? One current area is to use the word processing power of computers to stimulate interest in writing and to free children from some of the low level aspects of writing (such as correct spelling, punctuation and penmanship). For example, Sardamalia, Bereiter and Geolman (1982) propose that since the information processing capacity of young writers is limited, and since the mechanical and syntactic aspects of writing are not automatic, emphasis on correctly formed sentences results in poorer overall writing quality. The low level aspects of writing interfere with higher level planning. Evidence for this assertion includes the finding that when children are allowed to dictate their essays (which presumably frees them from some of the low level aspects of writing) they produce longer and higher quality essays as compared to writing.

Currently available word processing systems make revision much

easier and free the writer from some aspects of production (such as penmanship and spelling). However, word processors of the future will be even more helpful in stimulating high quality writing. For example, the "writer's workbench" (Macdonald, Frase, Gingrich, & Keenan, 1982) is an intelligent computer coach. It consists of a collection of programs which analyze written prose and make suggestions for revisions. The writer's workbench is actually in use at Bell Laboratories, with over 1,000 users. You can type your text into the computer, using a standard word processing system. Then, once you have finished your first draft, you can ask the programs from the writer's workbench to suggest revisions in your manuscript.

The writer's workbench consists of three major parts: a proofreader, a style analyzer, and an on-line English reference guide. The proofreader consists of the following programs:

spelling--lists all words that may be misspelled, and allows the user to specify any new words (such as jargon, proper names, and acronyms) to the list of acceptable words.

punctuation--lists cases where punctuation may be needed or where existing punctuation may be incorrect.

double words--lists all cases in which a word is repeated.

faulty phrasing--lists phrases which may not be coherent.

split infinitives--lists all instances of split infinitives.

An example of the output of the proofreading program is shown in Figure 5. As can be seen, the program points out possible errors as well as making suggestions for how to correct the errors.

The style analyzer consists of the following programs:

Figure 5. Output From A Proofreader Program

INPUT: Our report, "The Basic Fundamentals of Computatonal Complexity", is enclosed. Please send any recomended changes at your earliest convenience. thanks.

PROOFR OUTPUT:

************************** SPELLING **************************

Possible spelling errors in examplefile are:

Computatonal recomended

If any of these words are spelled correctly, later type
spelladd word1 word2 ... wordn
to have them added to your spelldict file.

************************** PUNCTUATION **************************

The punctuation in examplefile is first described.

2 double quotes and 0 single quotes
0 apostrophes
0 left parentheses and 0 right ones

The program next prints any sentence that it thinks is incorrectly punctuated and follows it by its correction.

line 1
OLD: Our report, "The Basic Fundamentals of Computatonal Complexity",
NEW: Our report, "The Basic Fundamentals of Computatonal Complexity,"
line 3
OLD: earliest convenience. thanks.
NEW: earliest convenience. Thanks.

For more information about punctuation rules, type:

punctrules

************************** DOUBLE WORDS **************************

For file examplefile:

No double words found

************************** WORD CHOICE **************************

Sentences with possibly wordy or misused phrases are listed next, followed by suggested revisions.

beginning line 1 examplefile
Our report, "The *[Basic Fundamentals]* of Computatonal Complexity", is enclosed.

beginning line 2 examplefile
Please send any recomended changes *[at your earliest convenience]*.

file examplefile: number of lines 3, number of phrases found 2

——————— Table of Substitutions ———————

PHRASE SUBSTITUTION

at your earliest convenience: use "soon" for " at your earliest convenience"
basic fundamentals: use "fundamentals" for " basic fundamentals"

************************** SPLIT INFINITIVES **************************

For file examplefile:

No split infinitives found

style--provides readability indices, measures of average word length and average sentence length, the percentage of verbs in the passive voice, the percentage of nouns that are nominalizations, the number of sentences that begin with expletives, and other such information.

prose--compares the style statistics listed above with some standard measures; if the text's measures are outside of the standards, the program prints an explanation of why the text may be hard to read and prints suggestions for how to correct the problem.

find--locates individual sentences that contain passive verbs, expletives, nominalizations, "to be" verb forms, and other potential problem sentences.

The on-line reference programs include information on the correct use of 300 commonly misused words and phrases, a computerized dictionary, and general information about the writer's workbench. Additional programs rate the words in the text for abstractness-concreteness, rate the paragraph organization, and detect possible instances of sexist language.

Other writer's helper systems include JOURNALISM, a proofreader that comments on the organization and style of news stories (Bishop, 1975), and CRES, a proofreader that identifies uncommon words, long sentences, and difficult phrases in NAVY documents (Kincaid, Aagard, O'Hara, & Cottrell, 1981).

Intelligent computer coaches for writing may [illegible] writers to develop more productive writing strategies. For examp[illegible] early drafts more attention can be devoted to the organization and goals of

the document, since proofreaders will detect lower level errors. In addition, writers are encouraged to engage in more extensive revision cycles, allowing for refinement of writing strategies. Unfortunately, there is very little empirical information concerning the effectiveness of writing coaches, but Macdonald et al. (1982) report that writers tend to like the programs.

Goldstein (1980) has developed a computer coach to teach general problem-solving strategies. For example, a student is asked to play a computer game that requires the use of strategic thinking. Throughout the game, the computer coach makes suggestions or observations about the strategy that the student is using. Goldstein (1980, p. 53) states that "the coach's function is to intervene occasionally in student-generated situations to discuss appropriate skills that might improve the student's play." Thus, an ultimate use of computers may be to expand the power of human strategic thinking. However, as Hayes and Flower (1980) and Goldstein (1980) have pointed out, successful computer coaches must be based on useful theories of human thinking (such as Newell & Simon, 1972). Again, the usefulness of a computer coach is tied to the underlying theory of cognitive processing.

Conclusion

We began with a pessimistic and an optimistic scenario for the role of computers in education. This paper then briefly explored examples of how computers can be used to help learners acquire semantic, procedural, and strategic knowledge. The major theme of this paper has been that the effective use of computer technology in schools is tied to the educational value of current theories of human learning and cognition. Another way to state this theme is to say

that the future of computer technology in schools depends on both the technological power of computers and the pedagogic usefulness of cognitive psychology.

A quarter of a century ago, American education was introduced to the technological innovation of teaching machines supported by a behaviorist psychology of learning. Today, schools are again being asked to participate in a technological revolution; however, the technological innovation involves computers, and the dominant psychology of learning is cognitive psychology. The realization of the optimistic scenario depends on our ability to extract what is useful from the cognitive psychology of human learning and cognition, and to creatively apply the information to the development of computer-based instructional mateials. Blindly using computers, without making use of what we now know about human learning and cognition, is likely to result in the realization of the pessimistic scenario.

References

Bishop, R.L. (1975). The JOURNALISM programs: Help for the weary writer. Creative Computing, 1, 28-30.

Brown, J.S., & Burton, R.R. (1978). Diagnostic models for procedural bugs in basic mathematical skills. Cognitive Science, 2, 155-192.

California State Department of Education. (1982). Proposed graduation requirements. Sacramento, California.

Collins, A. (1977). Processes in acquiring knowledge. In R.C. Anderson, R.J. Spiro, & W.E. Montague (Eds.), Schooling and the acquisition of knowledge. Hillsdale, NJ: Erlbaum.

Collins, A., & Stevens, A.L. (1982). Goals and strategies of inquiry teachers. In R. Glaser (Ed.), Advances in instructional psychology: Vol. 2. Hillsdale, NJ: Erlbaum.

Deringer, D.K., & Molnar, A. (1982). Key components for a national computer literacy program. In R.J. Seidel, R.E. Anderson, & B. Hunter (Eds.), Computer literacy. New York: Academic Press.

Fortune, R.F. (1982). The computer: Its impact on the classroom. California School Boards, 42, (3), 5-6.

Fuson, K.C. (1982). An analysis of the counting-on procedure in addition. In T.P. Carpenter, J.M. Moser, & T.A. Romberg (Eds.), Addition and subtraction: A cognitive perspective. Hillsdale, NJ: Erlbaum.

Goldstein, I. (1980). Developing a computational representation for problem solving skills. In D.T. Tuma & F. Reid (Eds.), Problem solving and education. Hillsdale, NJ: Erlbaum.

Groen, G., & Parman, J.M. (1972). A chronometric analysis of simple addition. Psychological Review, 79, 329-343.

Hayes, J.R. & Flower, L. (1980). Identifying the organization of writing processes. In L.W. Gregg & E.R. Steinberg (Eds.), Cognitive processes in writing. Hillsdale, NJ: Erlbaum.

Kincaid, J.P., Aagard, J.A., O'Hara, J.W., & Cottrell, L.K. (1981). Computer readability editing system. IEEE Transactions on Profession Communication, PC-24, 38-41.

Macdonald, N.H., Frase, L.T., Gingrich, P.S., & Keenan, S.A. 1982). The writer's workbench: Computer aide for text analysis. Educational Psychologist, 17, 172-179.

Mayer, R.E. (1981). The promise of cognitive psychology. San Francisco: Freeman.

National Council of Teachers of Mathematics. (1980). An agenda for action: Recommendations for mathematics in the 1980's. Reston, Virginia.

National Science Foundation. (1979). Science and engineering education for the 1980's and beyond. Washington, DC.

Newell, A., & Simon, H.A. (1972). Human problem solving. Englewood Cliffs, NJ: Prentice-Hall.

Nora, S., & Minc, A. (1980). The computer of society: A report to the President of France. Cambridge, MA: MIT Press.

Papert, S. (1980). Mindstorms. New York: Basic Books.

Scardamalia, M., Bereiter, C., & Goelman, H. (1982). The role of production factors in writing ability. In M. Nystrand (Ed.), What writers know. New York: Academic Press.

Seidel, R.J., Anderson, R.E., & Hunter, B. (1982). _Computer literacy._ New York: Academic Press.

Skinner, B.F. (1968). _The technology of teaching._ New York: Appleton-Century-Crofts.

Skinner, B.F. (1958). Teaching machines. _Science_, _128_, 969-977.

Instructional Software Design Principles

David Merrill

University of Southern California

Instructional design has been reborn as a cottage industry. People who have never before been engaged in the preparation of educational materials, and who are unfamiliar with the important work in the field, are currently designing instructional software and selling it through the mail. However, because so many software writers lack the necessary grounding in design principles, they are inadvertantly re-inventing the wheel, and doing a poor job of it.

Design Principles

Many instructional programs that I've reviewed are generality-rich and example-poor: they provide lots of definitions but not enough examples. There is a tendency to put a lecture or a textbook on the screen. In fact, I've been told by the representative of a major publishing company that their practice is simply to transfer their own workbooks to the computer monitor screen. This company publishes programs that are no more than the standard school workbooks. The correct approach is to represent each major idea with a rule, an example, and as much opportunity as possible for the student to practice the information. As simple an idea as that is, it is frequently violated, not only in computer-assisted instruction but in almost all instruction. A program should ask the students not to repeat the information that has been presented, but to apply it and demonstrate what they have learned.

Putting a textbook on the screen is a mistake. There is an economic justification for filling a printed page with text, but there

is no reason to solidly pack the computer screen with words. ~~Information presented in a block is hard to read. It is not organized~~ in a way that makes it easy to interpret. If the program does no more than duplicate the format of a textbook, the student might as well be given a traditional book. The special audio and video capabilities of the computer should be utilized as attention-focusing devices to tie the information together and emphasize important points.

The program should encourage active mental processing. The Skinnerian principle of overt response may or may not indicate that the student is actually thinking about the lesson, so a response alone does not constitute an adequate interaction between student and computer. It is necessary to anticipate a wide variety of student reponses in order to facilitate the student's active involvement. Also following this principle, constructed responses are better than multiple choice items. And before the student is required to practice the information given, the program should provide expository examples. Many students feel threatened by an instruction to go directly from reading a definition to coming up with an example of their own, and prefer to see the principle illustrated first.

The student should control the text output. Pacing the presentation can be an effective emphasis, but some means should remain for the student to recall information if necessary, to gc faster or slower as desired. Automatic scrolling, as provided by Pilot and other computer languages, can be irritating. A dynamic display that uses timing and stress to improve readability is far

better than a screen full of text. Animation is one good way to show relationships between different pieces of information.

Learner Input and Control

Keyboard anxiety is a problem for students, teachers, executives--people who regard typing as a low-level skill, never learned to type, and are consequently intimidated and threatened by the need to convey instructions to the computer by means of a typewriter-like keyboard. This problem will diminish in time, as more people recognize typing to be as basic a skill as handwriting. But at the moment there is a need to minimize typing. We need to use procedures other than finding letters on a keyboard that make it easier to go through a program. One possibility is the use of arrow keys to point to options on the menu. It's easy to use a pointer; you can point to the menu and touch a return key and accomplish what you want that way. Use arrows to go forward or backward and for page turning. Pointers can be used when responses don't have to be typed out. There are times when the purpose of the lesson involves having the student type out the words in full, but in the case of multiple choice selections it makes much more sense to use a pointer than to type letters. Using a pointer also minimizes the chance of accidentally making an incorrect entry.

An advisor function should be provided. We have to monitor the student's activity, not only in terms of whether his or her answers are right or wrong, but in terms of whether the student has seen enough examples and is progressing through the program without skipping important information. It's very easy to store data on

whether the students are paying attention and making consistent responses, or whether they have made so many errors that they need additional practice. This data can be used to advise the students on their progress and offer them alternatives.

Students are very good at selecting the number of examples they need before they understand a rule. There is no point in having a student who has already mastered an idea look at more and more examples. Nevertheless, there are students who abuse the privilege. Given the opportunity to leave one part of the program and advance, some students will move ahead before they are adequately prepared. To guard against this, the program can tell them that they have not yet done enough work in the previous section.

Learner control, then, has both advantages and disadvantages, but should always be provided for. Control of text output is certainly required; nothing should go past the students before they've had a chance to assimilate it, nor should a fast reader be chained to a slow presentation of text. The students should also be able to decide whether or not they want to ask for help; better students may be bored by a plodding step by step program. An escape function is also desirable, so that the student has the freedom to exit the program at will. It's a good idea to accompany the escape with more advice about where to begin when the student comes back to the program, especially if the escape is premature.

Programming Principles

Directions should be explicit and accessible. Some programmers forget that not everyone has their comprehensive knowledge of

conventional computer functions. Nothing should be assumed about the learner's computer knowledge and expertise. Options should be readily available to the student, either displayed on the screen or as a retrievable part of the program. However, in an attempt to help the novice, don't make it inconvenient for the experienced user. Provide a means whereby a learner who has the necessary skill can bypass the slower procedures that help an inexperienced user.

Always use the most natural procedures: escape keys should be used to exit the program, arrow keys should be used to move in different directions, the return or enter key should be used to enter commands. Make use of mnemonics so that menu commands employ the first letter of the word (or some other logical relationship) rather than the conventional listing of "A-B-C-D." Clever, arbitrary, unique definitions of key presses usually lead to confusion. Finally, structure the program. Like evaluation in education, everyone in computers talks about structured design but very few people do it. Structured design is critical to an educational program, especially when working with the limited space available on a small microcomputer with only 48K of memory. It's possible to design excellent programs within those constraints, but it demands a careful design that doesn't waste space.

Developing Domain-Specific Expert Systems

Bernard Gifford
Dean, Graduate School of Education
University of California, Berkeley

Instructional programs should operate at descending levels of sophistication to accomodate the student's level of understanding. This is the level at which it is appropriate to begin instruction. When the student begins to make mistakes which are not easily analyzed, he or she should be able to go to an error analysis routine. Diagnosis would then be made, not only of the type of error made, but of the reason for the error.

In order to develop a system which would display this kind of sensitivity to the learner's performance, a model of student problem-solving behavior is needed. It has only been within the last 20 years that psychologists have moved from research on instruction in a laboratory setting, and the development of generalized rules of knowledge acquisition, to the observation of students in the act of acquiring various kinds of knowledge. The shift in methodology, as well as the differentiation among types of knowledge, is due to the growth of cognitive science. There is still very little information on which to base a true instructional science that is domain-specific.

In addition to a model of how students acquire knowledge in a particular domain, a model is required of the successful problem-solving behavior that is desired as an outcome of instruction. Software is being developed which replicates the behavior of experts in various fields; in a medical program, for

example, very good doctors are asked how they go about diagnosis--what symptoms they look for, what questions they ask, what procedures they follow--and their responses form the basis of an expert system that can handle the large data bases that are available for diagnosing illnesses. An expert system in solving algebra problems would require a different set of skills, as would a system for geometry.

Most instructional programs being published do not provide adequate opportunity for interaction with the student; they are primarily drill and practice exercises. The student should have many chances to practice the appropriate problem-solving behavior.

Finally, a variety of fields of expertise should contribute to the design of educational programs. Knowledge of the problem-solving behavior of learners and experts (and of workable ways in which the two behaviors can be bridged) must be combined with a good understanding of the specific subject matter being taught. Unfortunately, it is possible to buy software that has obviously been written by someone who did not know enough about the subject, e.g., mathematics, to teach it effectively. Moreover, knowledge of instructional psychology and the principles of instructional design must be partnered by an understanding of the computer's capabilities. The juxtaposition of these diverse competencies is seldom achieved (and cannot realistically be expected in the immediate future) but is essential to the development of superior educational software

Full text will be available in monograph.

Some New (and Old) Directions for Computer Courseware

J.D. Fletcher

Center for Advanced Technology in Education

University of Oregon

In 1960 T.F. Gilbert wrote:

> If you don't have a gadget called a "teaching machine", don't get one. Don't buy one; don't borrow one; don't steal one. If you have such a gadget, get rid of it. Don't give it away, for someone else might use it...This is the most practical rule, based on empirical facts from considerable observation. If you begin with a device of any kind, you will try to develop the teaching program to fit that device. (p. 478. The emphasis is Gilbert's.)

This is a point of view with which many of us will sympathize. Educators who have mastered their craft through considerable investment of time and energy in learning how to use the traditional technologies of text, lectures, blackboards, and real-equipment laboratories have every right to be suspicious of new technology that threatens to revolutionize the hard-won techniques now at hand. Even programmers, initiates into the priesthood of computer technology, are occasionally elevated by computers to levels of frustration in which they are willing--and eager--to destroy thousands of dollars worth of equipment with their bare hands. Moreover, Gilbert is undoubtedly correct when he suggests that we may develop teaching

programs to fit the technology at hand. Of course we will, and to varying degrees we always have. To suggest that we should not pursue new technologies for this reason may not be so correct.

As Marshall McLuhan (1967) pointed out, every technology, to some extent, carries its own message. To ignore this message is to neglect the strengths of the technology. The tehnologies now becoming available will not only provide powerful new instructional tactics for presenting context, they will also make some content accessible that heretofore could not be taught in any practical setting. In the development of computer courseware it is possible to discern entirely new "functionalities" in instruction. As is true of most technological efforts, we have begun by trying to enhance the capability of our existing practice. We may end with new capabilities that change the nature of what we do in ways that are completely unanticipated. This could be the essence of the new computer revolution in schools. It is not just that we will have computers everywhere or that we will enhance our capabilities to instruct. We may also change our ideas about what instruction is. Not only will we get better at doing what we do now, but in a fundamental sense we may change what it is we want to do.

New Directions

It may be well to begin with a fable. This fable will already be familiar to some readers. Nevertheless, it seems sufficiently relevant to bear repeating. As the story goes, there once was a government "blue-ribbon" commisssion of instructional experts assembled to specify the ultimate in instructional technology. After

several days of meetings--suitably fueled by long lunches and accomodated by comfortable lodging--the experts came up with the following specifications for the new technology:

1. There should be no exotic power requirements. The technology should use ordinary household current, or be battery powered, solar powered, or require no power at all to operate.
2. It should be light and easily portable. One person should be able to transport it, and at best it would be carried in one hand.
3. There should be no complicated installation or environmental requirements. It should be easy to set up and use, it should operate in moderately extended temperature ranges, and it should be, as the miltary says, "ruggedized."
4. It should provide random access to a large amount of material.
5. It should be capable of displaying graphics, photographics, drawings, color, and high quality, easily read text.
6. It should be inexpensive, costing less than $50 a copy.

The commission report was received with great relief, for, as the perspicacious reader may realize, no research and development money was required to develop the technology. In fact, the technology already existed and had been in place for over five hundred years.

The appropriate technology was, of course, a book.

This is a fable for all of us in the business of applying new technology to instruction. We must come up with solutions that promise real innovations; in the case of instructional technology, they must be better than books. At the same time, some of our prototypes will be, like the horseless carriage, less efficient than what they are intended to replace.

Books are important beause, among other things, they are able to capture instructional content and make it inexpensively available to an unlimited audience. As Bunderson (1981) pointed out, computer technology is important because, among other things, it makes both the content <u>and</u> the interactions of great instruction inexpensively available to an unlimited audience. This promise has yet to be realized, but it seems almost inevitable. What we need to do is sift through all the prototypal development and find therein those embryonic techniques that promise to be better than books. It turns out that these techniques are neither easy to find nor trivial to develop. I will briefly examine them in the three areas of drill and practice, tutorial dialogue, and simulation.

Drill and Practice

"Drill and practice" is doubtless one of the more regrettable terms in instruction, evoking images of the classroom as a sweat shop and attracting the ire of those who want to use computers to create a rich and friendly learning environment for intellectual exploration and discovery in the classroom. Certainly it is now fashionable to deprecate drill and practice as a computer instruction technique, and

it has been so for the last five years. Papert (1980) cites drill and practice as an example of the QWERTY phenomenon. It turns out that because the mechanical keyboards of earlier times were unable to keep up with skilled typists--the keys would jam and otherwise misbehave if they were operated too quickly--typewriter keyboards were originally designed to slow down the key presses of skilled typists. The result was the QWERTY keyboard, named after the topmost row of letters. This keyboard is with us today despite our having removed all the mechanical obstacles to fast operation that results in the QWERTY design in the first place.

Papert's argument is that early applications of computers to instruction necessarily followed drill and practice formats partly because that is what classroom teachers would accept and partly because the computer technology of earlier times could support nothing else. This point of view is not entirely accurate, as can be seen in the design of curricula for the IBM 1500 System in [illegible] mid-1960's. The Stanford beginning reading program is a case in point. This curriculum, which was designed roughly in the period 1964-1966 and is described more fully by Fletcher (1979), encouraged children to build matrices using words and spelling patterns, to read and to be read stories (with illustrations), to record and play back messages, and to experiment with linguistic forms and variations.

Teacher acceptance was an issue somewhat separate from the content and approach of the curriculum--using computers to teach at all and taking away from classroom time to do it were the central concerns of the teachers. Nonetheless it is notable that when the

Stanford group moved to a less expensive machine configuration for presenting beginning reading instruction, the curriculum became more drill and practice in nature.

In any event, it seems past time to make a few arguments in favor of drill and practice. Is drill and practice an example of Papert's QWERTY phenomenon? The answer seems to be "no", partly because it works--drill and practice is still one of the most successful techniques we have in computer instruction--and partly because there is so much yet to be tried and developed in the drill and practice mode. Even if we assume drill and practice is limited to presentation of discrete items such as math facts or vocabulary items, there are at least three directions for curriculum development in drill and practice. These have to do with performance goals, optimal item selection, and optimal activity selection.

Performance Goals

We may best begin with trajectory theory. Basically this is a way of accounting for the progress, or trajectory, of individual students through a curriculum as a function of the amount of time they spend working in the curriculum. Figure 1 shows, perhaps more clearly, what trajectory theory is getting at. For individual students A, B, and C we try to predict and prescribe their grade placement on standardized paper and pencil tests based on the amount of time they spend on the computer curriculum. The interesting thing about trajectory theory is not just that it works, but that it has worked amazingly well in practice. In two published studies using trajectory theory (Suppes, Fletcher, & Zanotti, 1975 and 1976) the

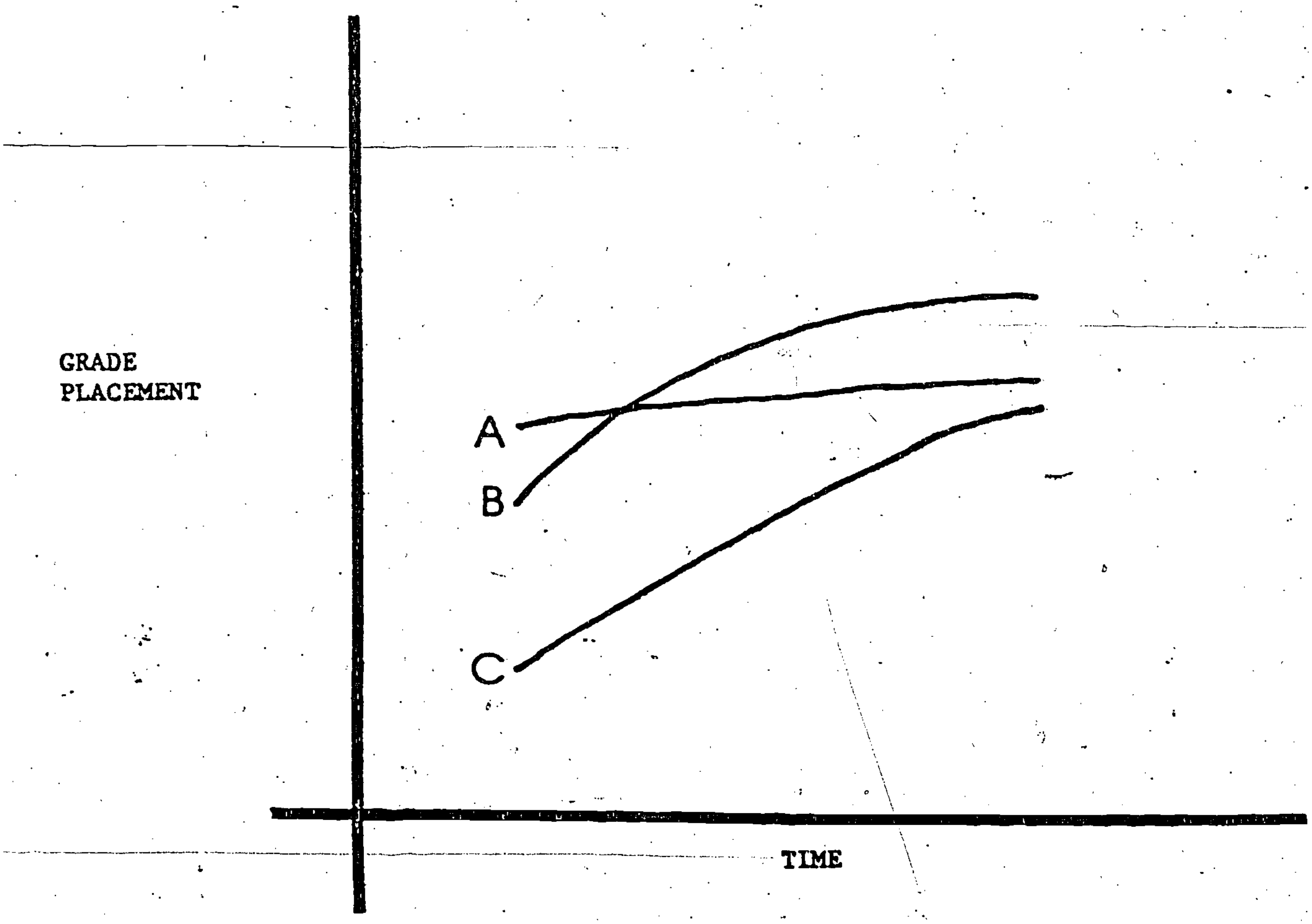


Figure 1. "Trajectories" of students through a curriculum.

standard error of estimated grade placement was in the range .04 - .06 of a grade placement. In other words, the estimates were off by less than a tenth of a grade placement for 90% of the cases. Again, these estimates were based solely on the amount of time the student spent on the computer and were independent of what was being done in the classroom. If we want to predict and control progress toward measured goals of achievement, trajectory theory may be one of the best techniques we have. It is worth emphasizing that although trajectory theory was developed for drill and practice, it may be applied to any form of instruction where we have closely watched and accurate measures of time on task, as we have in computer instruction.

There are still many questions to be answered about trajectory theory. Can it be applied to all subject matter? Can it be applied to methods of instruction other than drill and practice? Are there significant and important benefits to be gained from using classroom observations of time on task as well as computer time to predict and control progress? The list of questions could be continued. Trajectory theory is not a particularly new technique for computer curriculum, but it remains promising and worthy of further development.

Optimal Item Selection

Basically, an instructionally optimal solution is one that attempts to maximize some outcome, such as scores on an achievement test, subject to some constraints, such as total time on task, session length, and student ability. Optimal solutions are brought to use by control theory which in turn comes from operations research. It is a

well known and noted fact that operations researchers tend to attack problems by removing from them everything difficult to quantify or measure and building an imposing mathematical structure on what remains. In the current instances, the imposing mathematical structures remain, but some portion of what is difficult to quantify or measure can be supplied by mathematical models of learning and memory. The wherewithal for applying both these models and control theory to instruction in real time is provided by computers in the context of computer instruction.

The problem of optimal item selection in instruction was stated with mathematical clarity and rigor by Suppes (1964), but can be stated fairly simply in words: given a large number of items to be taught and a fixed time in which to teach them, what subset of items should be presented to an individual student at a given time in order to maximize his or her eventual achievement? The answer can be supplied by the above-mentioned quantitative models of learning and memory. Figure 2 presents a probability state-transition matrix of an appropriate sort based on General Forgetting Theory (Rumelhart, 1967; Paulson, 1973). This matrix shows what can happen to an item when it is presented to a student. As can be seen from the figure, the model of learning postulated is very simple. If an item is in the learned state, it stays there. If it is in the short-term state, it will either advance to the learned state or stay where it is. If it is in the unlearned state, it will advance to the short-term state or the learned state or remain unlearned. General Forgetting Theory is actually a little more sophisticated than this in that it accounts for

LEARNING STATE AT TIME $T+1$

LEARNING STATE AT TIME T

$$
\begin{array}{ll}
 & \begin{array}{ccc} \text{LEARNED} & \text{SHORT-TERM} & \text{UNLEARNED} \end{array} \\
\begin{array}{l} \text{LEARNED} \\ \text{SHORT-TERM} \\ \text{UNLEARNED} \end{array} & \begin{bmatrix} 1 & 0 & 0 \\ c & 1-c & 0 \\ a & b & 1-a-b \end{bmatrix}
\end{array}
$$

Figure 2. Probabilities of an item state transition when it is presented at time T.

probabilities of correct responding separate from the learning status of items and, notably, it postulates what happens to the learning status of an item when it is <u>not</u> presented. An optimal strategy for item selection based on General Forgetting Theory is, like all models of this sort, fairly simple in its view of human learning but fairly complex to imp ment. It could not be implemented by a book.

Studies orton (1973) for teaching spelling and by Laubsch (1970) for teaching foreign language vocabulary have shown approaches of this sort to be effective. They may even be dramatically effective, far more so than any other method for teaching large numbers of relatively independent items to students, but little work has been done in them since the mid-1970's. It seems to be a thread of research we have let slip through the cracks. There seems to be no real reason to drop it from our list of new directions for computer curriculum. Its promise for exceedingly efficient instruction remains.

Optimal Activity Selection

A few words may also be in order for optimal selection of activity. This problem most clearly emerges in the context of "strands" approaches to curriculum development. The strands approach, which was first described by Suppes (1967), calls for the apportioning of a computer curriculum into various content areas, or strands. For instance, a curriculum in reading comprehension might be divided up into vocabulary, literal comprehension, and interpretive comprehension strands. The problem, then, for a computer curriculum designer is to decide how much time students should spend in each strand or, to

state it a little more completely, how to control student progress in each strand so that each student's achievement over all strands is maximized at the end of some specified period of time. If progress in each strand is independent of progress in each of the others and if each of the strands contributes equally to the measure of achievement, then the solution is simple: we just pick the strand in which learning rate is greatest and allocate all the student's time to it. If, however, the situation resembled our reading comprehension example in which progress in one strand is interrelated with progress in the others, the situation is more complex. In reading, after all, a student with a poor vocabulary will not progress very far in literal or interpretive comprehension, yet the achievement measure of success for the curriculum will presumably be more concerned with comprehension than with vocabulary growth. Some sort of optimal mix of vocabulary development and work in comprehension will have to be devised for the student.

An appropriate optimal strategy (based on the Pontryagin maximum principle of control theory) for adjusting progress in interrelated strands was devised by Chant and Atkinson (1973) for the case of two strands. This strategy determines how much time a student should spend each day in each strand, depending on the student's learning rate in each strand and on how much he or she has progressed already in the strand. Extension of the strategy to curriculum environments with three or more strands was left by Chant and Atkinson as an exercise for the reader, but was described by the authors as being "straightforward". It very probably is, but it has not been done, or

at least it has not been published. Moreover. there have been no applications of this strategy to determine in practice how much it really buys in terms of student achievement relative to other approaches. In other words, here is another promising direction which we ha[illegible] just begun to explore. It cannot be implemented in a book, [illegible]re needs t[illegible] done.

Most [illegible] [illegible]cal psychologists reading the above discussion of drill and practice will find it difficult to suppress dark uncomplimentary mutterings about "1960's psychology". There are cycles in research, as in most things. In this dimension, we seem to oscillate between attacking small, tightly constrained, and fairly uninteresting problems over which we exercise a great deal of control, and attacking very large, sloppy, and interesting problems over which we can exert very little control. As may be evident from the above discussion and from reviews by Atkinson and Paulson (1972) and Fletcher (1975), drill and practice emphasizes the former. Nonetheless, it should also be evident that drill and practice is not just a matter of throwing items at students who are treated in some assembly line fashion. There are deep, educationally significant, and scientifically credible issues yet to be settled concerning drill and practice. Finally, it should be evident that despite the early strong results we have had from drill and practice, much more could be done to fully realize the promise of this approach.

As far as the oscillation between tightly controlled, less interesting problems and poorly controlled but much more interesting problems is concerned, it appears that current research in psychology,

applied psychology, and instruction emphasizes the latter. This trend is especially apparent in current attempts to build tutorial dialogue systems. Nowhere is the attempt to automate single tutor/single student dialogue more evident. This is the line of development to turn to next.

Tutorial Dialogues

Before diving into the area of tutorial dialogues, a few comments on the automation of programmed textbooks may be in order. Most commentators on tutorial dialogue approaches include in this category the intrinsic programming techniques of Crowder (1959) that appear so frequently in commercially available computer instruction materials. Basically this approach uses the computer as a very large and sometimes very intricately programmed textbook. This is an approach that could be pursued in a book, although the book might have to be carried around in a wheelbarrow. Nonetheless, this approach appears to concern application of book and text technology rather than computer technology to instruction. It remains one of the most common, easily produced, and frequently implemented approaches, and it is best supported by authoring languages for computer instruction. The development of authoring languages such as PILOT, TUTOR, WISE, PLANIT, etc., all seem to have intrinsic programming in mind since this is the approach most easily taken when one uses these languages.

We tend not to publish our unsuccessful attempts at computer instruction, among other things, but there seems to be an underground consensus among those in the business that these intrinsic programming aproaches do not work very well. What appear to be intuitively

obvious and correct procedures for assessing student knowledge, deciding when to branch, and providing remedial and accelerated material turn out to be relatively ineffectual in the light of student performance data. The determined reader is welcome to peruse Fletcher and Beard (1973) as an example of unpublished--and unsuccessful--work of this sort. In any case, this section does not concern the automation of programmed textbooks.

This section is conerned with the development of intelligent instructional systems as a new direction for computer instruction. This approach is a direct attempt to imbue computers with the qualities of expert human tutors. This line of development grew out of early concern with just how long it took, and how expensive it was, to generate items for computer presentation. Early estimates of the amount of time required to produce one hour of computer instruction ranged from 77 to 714 hours on PLATO, 200-400 hours on TICCIT, and around 475 hours for the IBM 1500 Instructional System (Orlansky & String, 1979). One solution to this problem was sought by those who noticed that the process of preparing items for computer presentation was boring, repetitious, and dull--in other words, a perfect job for computers. The resulting solution took the form of programs that would generate items for students (e.g. Koffman & Blount, 1974) and was called Generative Computer-Assisted Instruction, although what we now mean by generative computer instruction is a little more sophisticated. In any event, it occurred to early observers of the scene that since we were trying to use computers to mimic the item generation capabilities of expert human tutors, why not use computers

to mimic all the capabilities of human tutors? Thus was born the notion of computerized tutorial dialogue.

The development of computerized tutorial dialogues involves the application of artificial intelligence techniques to computer instruction, resulting in the information structure oriented (ISO) approaches discussed and advocated by Carbonell (1970). Carbonell contrasted these approaches with ad hoc frame oriented (AFO) approaches based on techniques of programmed instruction. Carbonnell pointed out that, unlike AFO approaches, ISO approaches can be used to develop instructional systems that answer questions not specifically anticipated by the instruction designers, construct appropriate questions on given topics, and carry on a "mixed-initiative" dialogue in which either the student or the computer can introduce a response, topic, or idea in a free and comfortable subset of English. This may sound like programming a computer to be an expert tutor, and it is meant to.

This approach is in the mainstream of current developments in cognitive psychology which have taught us--or reminded us--that perception and learning are overwhelmingly constructive processes (cf. Resnick, 1983). In perception we do not collect bits of information from the "outside world" and paste them up on perceptual templates, and in instruction we are not writing information on blank slates in students' heads. Instead, we are dealing with active and very rich simulations of the world which students must create in order to perceive or learn. It is analysis by synthesis with a vengeance, and what gets transmitted in communication and instruction are not

bits of information or knowledge but cues that may or may not be used to adjust the simulations being built up by students. The attempt in tutorial dialogue approaches is to deal directly with these simulations in ways that no drill and practice program--and no book--can.

Computers are both very good at this and very bad. Consider the following sentence:

The man the dog the girl owned bit died.

This is a difficult sentence for us to parse. We quickly become entangled in its syntactic nestings. Human chauvinism leads us to assume that since the sentence is difficult for us to parse, it is impossible for a machine. Yet a computer could quickly discern, after diving into its recursive routines for processing nested constructions, that there was a dog that was owned by a girl, that the dog bit a man, and that the man subsequently died.

Here is another example:

The man kicked the ball kicked the ball.

This is a perfectly grammatical sentence, as any self-respecting machine would discover after reversing an English transformational rule for deleting function words and determining that a man to whom a ball was kicked, kicked the ball back. In both these examples, a computer is less likely than we are to be confused or distracted, and its ability to process these two examples illustrates real intellectual ability. "Artificial intelligence" is, after all, a poor name for the business of making computers intelligent. Intelligence, or intellectual ability, is really what the ıd is all about. That

theories of intelligence are tested by algorithmization and putting them on computers is merely an issue of methodology, albeit a central one; there is nothing artificial about the capabilities targeted by this work.

Next we might consider the following example, taken from Donald Norman (1973):

What was Charles Dickens's telephone number?

A knowledgeable program would search the attributes it had associated with Charles Dickens looking for a telephone number, and/or it would search its list of telephone numbers looking for one associated with Charles Dickens. Presumably, no telephone number will be found for Charles Dickens, and this fact will simply be reported. As most human information processors must be aware, there is a more intelligent answer to this query that completely avoids searching lists of telephone numbers and/or attributes. To reply that telephones were not used until after Dickens's lifetime does require some knowledge, but this information could easily be stored by the computer. The real problem is to devise a computer program that knows when this information is relevant.

These examples also illustrate the distinction drawn by John Brown between "intelligent" computer instruction and what might be called "knowledgeable" computer instruction. In this light we might consider the following student/computer protocol taken from Brown's SOPHIE system (Brown, Burton, & Bell, 1974). In this example, the student is completing a series of tests on a computer simulated piece of electronic equipment (a regulated power supply) with a fault

generated by the computer. Student reponses are underlined.

WHAT IS THE BSE EMITTER VOLTAGE OF THE VOLTAGE LIMITING TRANSITOR?

BASE

TRANSISTOR

THE BASE EMITTER VOLTAGE OF Q5 IS -.037 MILLIVOLTS.

The spelling corrector has detected and corrected two misspellings. The system must also know that Q5 is the voltage limiting transistor.

WHAT ABOUT THE CURRENT LIMITING TRANSISTOR

THE BASE EMITTER VOLTAGE OF Q6 IS .349 VOLTS.

The meaning of this question is completely dependent on the context established by the prior sentence. The system uses this context to determine that the student is interested not only in the voltage measurements but in the voltage across a particular junction.

WHAT SHOULD IT BE?

IN A WORKING CIRCUIT THE BASE EMITTER VOLTAGE OF Q6 IS .593 VOLTS.

Again the context determines the meaning of the question. Moreover, the system must infer that "should" implies what the measurement would be in a properly working device.

REPLACE Q6

I AM GOING TO ASK YOU SOME QUESTIONS ABOUT HOW Q6 IS FAULTED. ARE ANY JUNCTIONS SHORTED?

At this point SOPHIE knows that the student is barking up the wrong tree. Rather than simply replacing the transistor as requested, SOPHIE begins to query the student about what he or she thinks is wrong. If the student gets at least something right, the system says what was actually wrong and goes ahead and replaces the faulty part.

Clearly, SOPHIE has progressed beyond knowledgeable syntactic and semantic processing of student inquiries and is behaving in an "intelligent" manner. The system tries to create a reactive environment in which the student learns by actively trying out ideas in interaction with a simulation program. However, the program does more than simply simulate the information to be transmitted; it provides for tutorial feedback and, in effect, for a one-to-one relationship with an "articulate expert" problem solver who helps the student create, experiment with, and debug his or her own ideas.

Several reviews of this area have appeared, notable among which are discussions by Peele and Riseman (1975), Sleeman and Brown (1982), Barr and Feigenbaum (1982), and Fletcher (1984). Fletcher references about 16 of these tutorial dialogue systems that have been or are being developed. Carbonell's SCHOLAR (1970) and Brown's SOPHIE (Brown, Burton, & Bell, 1974) were seminal systems in the development of tutorial dialogues. The two premier systems currently seem to be GUIDON (Clancey, 1979) and Steamer (Williams, Holland, and Stevens, 1981).

GUIDON serves as a physician's consultant for the student, who plays the role of the physician, in diagnosing infectious diseases. GUIDON focuses directly on the problems a subject matter expert faces in making his or her expertise, understanding, and heuristics accessible to students. GUIDON takes account of students' knowledge and interests in choosing what to present, it incorporates a knowledge base that is augmented to better organize and explain the subject matter to the student, and its teaching expertise is represented

explicitly and modularly so that it can be modified for different research designs. GUIDON both "knows" the subject matter and can explain to the student the paths it uses to reach a diagnosis just as an expert tutor does.

Steamer is a computer-based system being developed by the Navy to provide instruction in steam propulsion engineering. It links a very complicated and highly abstract, quantitative (mathematical) model of a ship's steam propulsion system to high quality visual (graphics) presentations of the underlying model. The student is thereby able to manipulate the underlying abstract model through the graphics interface and to see in computer graphics presentations how the effects of these manipulations would be propagated throughout the ship's steam propulsion system. Additionally, Steamer uses the student's manipulation to better model his or her understanding of steam and to extend, correct, and deepen that understanding.

At this point, we may all wonder if we are going to see tutorial dialogue systems of this sort in our classrooms in the near future. About a year ago one of the major figures in the tutorial dialogue world passed through Oregon State University leaving the following quote in his wake: "It's amazing what you can do when you only have two megabytes of memory."

To those of us used to working with 32K and 64K byte personal computers, the notion of 128K bytes seems like Nirvana. Two million bytes is beyond all imagining, and this is apparently the low end for someone working with tutorial dialogues. The point is that the computational requirements for tutorial dialogue systems are very

large. A single user system sufficiently powerful for delivery but not development of tutorial dialgues might be purchased today for about $20,000. In ten years the picture will change completely, and for this reason the development of tutorial dialogue systems should now be pursued vigorously on large machines.

Somewhere among all the new directions for computer courseware a major breakthrough will occur. Tutorial dialogues appear to be a likely area for this breakthrough. This direction represents an approach that is both evolutionary and revolutionary. That is to say, we can expect it to help us accomplish what we want to do now and to alter in very fundamental ways our understanding of what instruction should be. In any event, tutorial dialogues could not be implemented without computers, and their development is limited by the current state of the art in both computer hardware and software. It is often said that hardware and software developments are far in advance of our capabilities to use them in instruction. In the case of tutorial dialogues, this is not true. We are simultaneously developing and capitalizing on the state of the art in computer hardware and software technology.

Much still needs to be done. We need to learn how to represent imperfectly understood and poorly described knowledge domains and to reduce the costs of creating knowledge domains. Better natural language processing must be developed, techniques for modeling learners must become far more sophisticated, and our understanding of what master tutors and teachers do must be greatly enhanced. We need to learn how to interface computer tutorial dialogues with the

practice of classroom teachers. However, these issues only indicate that breakthroughs in this area will occur perhaps later rather than sooner. The promise of tutorial dialogues for improving instruction remains.

This promise is particularly evident when we review efforts to join tutorial dialogue techniques with simulation, the topic of the next section. In fact, we have already skirted these shoals very closely. After all, the student troubleshoots a simulated power supply in SOPHIE, diagnoses an ailing simulated patient in GUIDON, and operates a simulated steam propulsion system in Steamer. It may be past time to turn to the area of simulation in instruction.

Simulation

The currently strong and growing interest in simulation used for education is far overshadowed by the interest in and support for simulation used in training, specifically military and industrial training. Most readers will be familar with the long history and use of multi-million dollar aircraft simulators--some costing more than the aircraft they simulate--by the military and by aircraft manufacturers for pilot training. Twenty years ago, if one mentioned the use of simulators in instruction the reference would be to aircraft simulators and probably nothing else. The advent of computer technology has permanently altered this state of affairs.

Because current simulators are based on programmable computers, they need not be single purpose, representing only a single system such as the cockpit of an F-14 fighter aircraft. Instead, a wide range of related systems can be simulated for the purposes of training

individuals who must learn to operate and maintain them. The Navy's Generalized Maintenance Trainer/Simulator (Rigney, Towne, King, & Moran, 1978) is a case in point. The GMTS can be used to simulate any device in which signal paths and their relationships to controls, indicators, and test points can be defined. So far the GMTS has demonstrated its versatility by being used to teach techniques to maintain both a radar repeater and a UHF communications systems.

Again because current simulators are based on programmable computers, they can be much smaller and less expensive than they were originally. Simulators too are benefitting from the micro-electronic revolution. The idea of "suitcase simulators" abounds in today's military. MITIPAC (Rigney & Towne, 1977), for instance, took the GMTS and shrunk it down via micro-electronics to fit into a suitcase-sized package which provides a true job site training capability. MITIPAC can now be transported to locations where military jobs are actually performed--in the field, on ships, on flight lines--and tailored to the specific jobs at hand. Many simulators have been built, tried, and evaluated in training, as Orlansky and String showed for training aircraft pilots (1977) and for training maintenance technicians (1981). In this sense, simulation is an established and proven technique for instruction. However, development of simulation for instruction is far from finished. The field is particularly fortunate in that promising and dramatic new "functionalities" now exist. Three of these new functionalities are interactive movies, surrogate

travel, and spatial data management. All three of these use computer-controlled videodiscs.

Interactive Movies

Interactive movies attempt to translate movie viewing into an active, participatory process. In effect, the viewer becomes the director and controls many features of the movie. Feature controls available to the viewer are the following:

1. Perspective. The movie can be seen from different directions. In effect, the viewer can "walk around" ongoing action in the movie or view it from above or below.
2. Detail. The viewer can "zoom in" to see selected, detailed aspects of the ongoing action or can "back off" to gain more perspective on the action and simultaneous activity elsewhere.
3. Level of instruction. In some cases, the ongoing action may be too rich in detail or it may include too much irrelevant detail. The viewer can hear or see more or less about the ongoing process by so instructing an interactive movie system.
4. Level of abstraction. In some instances the viewer may wish to see the process being described in an entirely different form. For example, the viewer might choose to see an animated line drawing of an engine's operation to get a clearer understanding of what is going on. In some cases, elements shown in the line drawings may be invisible in the ongoing action, e.g., electrons or force fields.

5. Speed. Viewers can see the ongoing action at a wide range of speeds, including reverse action and still frame.
6. Plot. Viewers can change the plot to see the results of different decisions made at selected times during the movie.

Surrogate Travel

Surrogate travel forms a new approach to locale familiarization and low cost instruction. In surrogate travel, images organized into video segments showing discontinuous motion along a large number of paths in an area are stored on videodisc. Under microprocessor control, the student accesses different sections of the videodisc, simulating movement over the selected path.

The student sees with photographic realism the area of interest, for instance, a city street or a hallway in a building. The student can then choose both the path and the speed of advance through the area using simple controls, usually a joystick. To go forward the student pushes forward on the joystick; to make a left turn the student pushes the joystick to the left; to go faster the student pushes the joystick harder, and so on.

The videodisc frames the viewer sees originate as filmed views of what one would actually see in the area. To allow coverage of very large areas, the frames are taken at periodic intervals that may range from every foot inside a building, to every ten feet down a city street, to hundreds of feet in a large open area, e.g., a harbor. Coverage of very small areas is also of interest. In microtravel, which is a combination of surrogate travel and interactive movies, travel is possible where humans could could never go: inside watches

while they are running, inside living organisms, etc.

The rate of frame playback, which is the number of times each video frame is displayed before the next frame is shown, determines the apparent speed of travel. Free choice in what routes may be taken is obtained by filming all possible paths in the area as well as all possible turns through all intersections. To some extent this a time consuming and expensive technology, but it has become relatively efficient because of the design of special equipment and procedures for doing the filming.

Demonstrations of this technology have been developed for building interiors (National Gallery of Art), a small town (Aspen, Colorado), an industrial facility (nuclear power plant), and San Francisco Harbor. Plans are underway to produce a prototype video map library of broader scope for selected areas worldwide.

Spatial Data Management

Basically, spatial data storage and retrieval of information is the method of loci transformed to a video or computer graphics format. The information is stored and retrieved through its association with already familiar geographic terrain.

Suppose, for instance, a student wanted to study the musical environment in which Ralph Vaughan Williams wrote his "Concerto for Tuba and Orchestra". In an ordinary data retrieval system the student will type in a complicated set of Boolean expressions--or English phrases standing for Boolean expressions--and will receive in return only textual information about the topic. Relevant information closely related to the information successfully retrieved will not

appear unless the student starts from the top again with a new set of Boolean expressions. In a spatially organized data system, the underlying geography will be familiar to the student, for instance the school campus. The student may then "fly" to the music department (or library, concert hall, professor's office, etc.) and look for a tuba (or an orchestra, music library, portrait of the composer, etc.). Upon finding a tuba or other relevant cue, the student can "zoom" into it, still using his single joystick contol, select the concerto by name (or by hearing it, seeing the score, seeing the composer, etc.) and then hear, see, and read more information about it all retrieved through visually oriented associations.

In this way, spatial data management acts as an electronic library that gives students and instructors access to a wide assortment of multi-source and multi-media information whose components are associated in a natural and easily accessible manner. Instructors can access the system to create and/or assemble their own information spaces to be explored later by their students or subsequently present these materials to large audiences in single locations using large screen television projection or to multiple locations though cable distribution systems. Students can independently use the system for individualized instruction by working though previously designed information spaces, by browsing on their own, or by creating their own data spaces. When students and instructors are in remote locations, offsite instruction can be facilitated by linking two or more systems together using regular telephone lines. In this manner, a student or instructor can "fly"

the other to a topic of interest, sharing at geographically remote sites a large, visually oriented library of information.

Two points are worth noting about these new directions for simulation applied to instruction. First, they cannot be implemented in a book. Second, the application of these new directions for simulation-based computer instruction in education is just beginning. One can easily imagine application of this technology to science education. Perhaps a few words on this subject are in order.

The best way to learn science is by doing it. The excitement, mystery, frustrations, and triumphs of science are only dimly revealed by the usual fare of introductory science course. It would be far better for students, especially introductory students, to approach science with freedom to indulge their curiosity, form and re-form their own hypotheses, design and perform their own experiments, and build their own models and theories to explain natural phenomena. Unless there are drastic shifts in national funding policies for science education, this essential scientific experience will be prohibitively expensive to provide. The result is that students--especially elementary and junior high school students--are "turned off" by science at a time when our industrial and academic need for scientists, engineers, and technologists is acute and increasing.

What is needed in science education is something that has the impact of video gaming, but at the same time possesses substantial pedagogical power. One way to accomplish this is to provide simulated scientific experiences to students. Good simulations are exciting,

compelling, and teach effectively by providing an environment in which learners must live with their decisions. Simulated experiences need not replace existing laboratory and field exercises, but they may expand and supplement them. Moreover, simulated experiences may be superior to real experiences in at least four ways. First, and primarily, simulation can be economical. Use of simulation should reduce the need for laboratory equipment and its maintenance, laboratory supplies, and travel costs for field experience. Second, simulation can make relevant phenomena more readily visible in two ways. In one way it can make the invisible visible. For instance, the flow of ions can be seen more clearly and simply under simulated conditions than under real conditions. In another way, simulation may increase the visibility of a phenomenon by separating it from a confusing or chaotic background. One can see the conceptual forest without getting lost in the procedural trees. Third, simulation allows reproducibility. Students can replay over and over chains of events that they could not otherwise observe repeatedly. Fourth, simulated ecxperience is often safer than the real thing. Airplanes can be crashed, poisons can be ingested, and laboratories can be exploded with impunity in simulated environments.

Two sorts of relevant scientific experience that lend themselves readily to simulation are field study and laboratory experimentation. These two kinds of experience could be provided using the new functionalities described above. These functionalities could be used to build video field trips and simulated laboratories.

In the field, the student sees the total ecological view. He/she

sees the overall landscape, the terrain, the populations of organisms, and individual samples of interest in their special areas. In sciences such as biology, geology, paleontology, archeology, and even astronomy, substantial learning and appreciation can be achieved by travel to locations that are difficult to access under the best of conditions. However, field trips are treated as an instructional frill. After all, the trips are made rarely and locally (they depend for success on what is serendipitously nearby); they emphasize only the group (individuals do not have an opportunity to do the science on their own); and most of the administrative effort centers on getting to the field and getting back, not on the field experience itself. As a result, even short, local field trips are being cancelled by schools because their cost in time and fuel is not balanced by their educational return. Surrogate travel removes the major objections to field experience and offers to each student a broadened opportunity to experience scientific phenomena in their natural, ecological context.

Students interested in, say, the biology of deserts could visit the Gobi in the morning, the Sahara around noon, and the Sonoran in the afternoon. They could travel around in each habitat locating, identifying, and "gathering" samples in roughly the same way, and for the same purposes, as a trained scientist. Panning and zooming through the full range of habitats could develop in students many of the same intuitions and understandings of environmental, geographic, and climatic contexts that an experienced scientist gains from actual travel.

Back in school, laboratories provide a problem solving

environment where students interact, observe processes, and are stimulated to synthesize concepts as part of their learning. However, many schools are eliminating laboratories from their science courses, not because they are not useful learning experiences, but because of the cost of obtaining, maintaining, and supporting specimens, samples, and laboratory equipment. Interactive movies and spatial data management allow us to simulate laboratory experiences without the high cost and effort that is normally involved under the present pattern.

Students can create, store, and retrieve information from mammoth data banks using spatial data management. One can imagine high school students organizing an entire archaeological excavation or geological survey using spatial data techniques. One can also imagine elementary school students setting up and running high-energy particle physics experiments through interactive movies with plot control. Students would also have full use of the latest in telescopes, microscopes, and even endoscopes through computer-based simulation.

Finally, laboratory and field experiences can be linked so that hypotheses developed in the laboratory would be tested by return "travel" to the correct habitat, "collection" of data or specimens, and return to the laboratory for testing and verification. In this way, the excitement, frustrations, and triumphs of scientific experiences would become accessible to students.

In the above, simulation was presented as a new direction that is finding its way into computer instruction, but it is interesting to note that the history of computer instruction is exactly the reverse.

The first use of computers to teach grew out of a computer based system that was primarily intended for simulation of real world experiences. This was the Air Force's SAGE (Semi-Automatic Ground Environment) system which was built in the late 1950's to train Air Force personnel in the techniques and tactics of air defense (Rowell & Streich, 1964; Parsons, 1972). Computers in SAGE were inititally used to simulate equipment, mostly radar, to which ground based air defense personnel were to make appropriate reactions. However, as time progressed, the SAGE computers began to be used to present training in a more general-purpose fashion.

The University of Illinois's PLATO (Programmed Logic for Automatic Teaching Operations) was probably the first computer system built specifically for computer instruction. Interestingly, it too was first supported solely by the military--in this case by the Army Signal Corps, the Office of Naval Research, and the Air Force Office of Scientific Research (Bitzer, Braunfeld, & Lichtenberger, 1962). Initially PLATO was used as a sort of "book with feedback" following the suggestion of Chalmers Sherwin, and few who saw early demonstrations of PLATO in the late 1960's were able to escape its "fruit fly" demonstration. This was a simulated biology laboratory showing in high quality graphics successive generations of fruit flies as they illustrated a model of genetics. This type of simulation in computer instruction is still in use.

The focus in this section is on new techniques for simulation, three of which are listed above. These three have been discussed in a little more detail by Bolt (1979) and by Levin and Fletcher (1981).

Other techniques may well be on the way. We have barely begun to explore the instructional possibilities of natural language processing (as opposed to computer language processing), voice output, voice input, computer-generated imagery (which may obviate some of the need for videodisc storage), and psychoneurological monitoring. New functionalities for these capabilities will doubtless be developed. However, it should be emphasized that this process of discovery is at least as demanding of time, resources, and ingenuity as the development of the computational capabilities themselves. Swamping schools with hardware and computer capabilities and then expecting instructional functionalities to flow spontaneously in their wake is simply wrong. The process will continue to require support, encouragement, resources, and time.

Final Word

It is wrong to inundate our educational institutions with new technologies without insisting that they do at least something to help us through the day. It is also wrong to hold off all investment in new technologies because they may affect what it is we want to do. The correct approach seems to be somewhere in the middle. No one envisioned teleconferencing when the telephone was invented, no one imagined our current interstate highway transportation system when the horseless carriage came along, and steam engines languished for 30 years pumping water out of coal mines before someone began to think seriously of their possibilities for self-locomotion. We have benefitted from the introduction of these devices into our lives just as we have suffered from them. We must give the new technologies

their place if we are to improve our instructional practice as the Gardner Commission said we must. At least in the case of computers, we are in a position to insist that they be of some immediate practical value along the way. This is a fortunate position to be in, and we should capitalize on it. Computers can help meet goals and solve current problems of schools and school districts at the same time they are helping to advance the craft of instruction. We can and should expect them to do both.

In short, computers will help us better perform the business of instruction as we envision it today. They will also broaden our horizons. They will change and expand our ideas about what instruction is and what it must do. Their challenge to us as educators is as serious as their promise. We should rise to the occasion.

References

Atkinson, R.C., & Paulson, J.A. (1972). An approach to the psychology of instruction. Psychological Bulletin, 78, 49-61.

Barr, A., & Feigenbaum, E.A. (1982). Handbook of artificial intelligence: Vol. 2. Los Altos, CA: William Kaufman.

Bitzer, D.L., Braunfeld, P.G., & Lichtenberger, W.W. (1962). PLATO II: A multiple-student, computer-controlled, automatic teaching device. In J.E. Coulson (Ed.), Programmed learning and computer-based instruction. New York: John Wiley and Sons.

Bolt, R.A. (1979). Spatial data management. Cambridge, MA: Massachusetts Institute of Technology.

Brown, J.S., Burton, R.R., & Bell, A.B. (1974). SOPHIE: A sophisticated instructional environment for teaching electronic troubleshooting (An example of AI in CAI) (BBN Report 2790). Cambridge, MA: Bolt Beranek and Newman, Inc.

Bunderson, C.V. (1981). Courseware. In H.F. O'Neil, Jr. (Ed.), Computer-based instruction: A state-of-the-art assessment. New York: Academic Press.

Carbonell, J.R. (1970). AI in CAI: An artificial intelligence approach to computer-assisted instruction. IEEE Transactions on Man-Machine Systems, 11, 190-202.

Chant, V.G., & Atkinson, R.C. (1973). Optimal allocation of instructional effort to interrelated learning strands. Journal of Mathematical Psychology, 10, 1-25.

Clancey, W.J. (1979). Tutoring rules for guiding a case method dialogue. International Journal of Man Machine Studies, 11, 25-49.

Crowder, N.A. (1959). Automated tutoring by means of intrinsic programming. In E.H. Galanter (Ed.), Automatic teaching: The state of the art. New York: John Wiley and Sons.

Fletcher, J.D. (1975). Models of the learner in computer-assisted instruction. Journal of Computer-Based Instruction, 1, 118-126.

Fletcher, J.D. (1979). Computer-assisted instruction in beginning reading: The Stanford projects. In L.B. Resnick & P.A. Weaver (Eds.), Theory and Practice of Early Reading: Vol. 2. Hillsdale, NJ: Lawrence Erlbaum Associates.

Fletcher, J.D. (in press). Intelligent instructional systems in training. In S.A. Andriole (Ed.), Applications in artificial intelligence. New York: Petrocelli Books.

Fletcher, J.D., & Beard, M.H. (1973). Computer-assisted instruction in language arts for hearing-impaired students (Technical Report No. 215). Stanford, CA: Institute for Mathematical Studies in the Social Sciences, Stanford University.

Gilbert, T.F. (1960). On the relevance of laboratory investigation of learning to self-instructional programming. In A.A. Lumsdaine & R. Glaser (Eds.), Teaching machines and programmed learning: A source book. Washington, DC: National Educational Association.

Koffman, E.B., & Blount, S.E. (1974). A modular system of generative CAI in machine-language programming. IEEE Transactions on Systems, Man, and Cybernetics, 4, 335-343.

Laubsch, J.H. (1970). Optimal item allocation in computer-assisted instruction. IAG Journal, 3, 295-311.

Levin, S., & Fletcher, J.D. (1981). Videodisc technology: An approach to the design of training devices. In J. Moraal & K.F. , Manned systems design: Methods, equipment, and applications. New York: Plenum Press.

Lorton, P.V. (1973). Computer-based instruction in spelling: An investigation of optimal strategies for presenting instructional material. Doctoral dissertation, Stanford University, Stanford, California.

McLuhan, H.M., & Fiore, Q. (1967). The medium is the message. New York: Random House.

Orlansky, J. & String, J. (1981). Cost-effectiveness of maintenance simulators for military training (IDA Paper P-1568). Arlington, VA: Institute for Defense Analyses.

Papert, S. (1980). Mindstorms. New York: Basic Books.

Parsons, H.M. (1972). Man-machine system experiments. Baltimore, MD: Johns Hopkins Press.

Paulson, J.A. (1973). An evaluation of instructional strategies in a simple learning situation. Doctoral dissertation and Technical Report No. 209. Institute for Mathematical Studies in the Social Sciences, Stanford University, Stanford, California.

Peelle, H.A., & Riseman, E.M. (1975). Four faces of HAL: A framework for using artificial intelligence techniques in computer-assisted instruction. IEEE Transactions on Systems, Man, and Cybernetics, 5, 375-380.

Resnick, L.B. (1983). Mathematics and science learning: A new conception. Science, 220, 477-478.

Rigney, J.W., & Towne, D.M. (1977). Some concepts, devices, and procedures for improving technical training and on-the-job performance of maintenance (Technical Report 76-C-0023-1). Orlando, FL: Naval Training Equipment Center.

Rigney, J.W., Towne, D.M., King, C.A., & Moran, P.J. (1978). Field evaluation of the generalized maintenance trainer-simulator: I. Fleet communications system (Technical Report No. 89). Los Angeles, CA: Behavioral Technology Laboratories, University of Southern California.

Rowell, J.T., & Streich, E.R. (1964). The SAGE system training program for the Air Defense Command. Human Factors, 6, 537-548.

Rumelhart, D.E. (1967). The effects of interpresentation intervals on performance in a continuous paired-associate task (Technical Report No. 116). Stanford, CA: Institute for Mathematical Studies in the Social Sciences, Stanford University.

Sleeman, D., & Brown, J.S. (Eds.). (1982). Intelligent tutoring systems. New York: Academic Press.

Suppes, P. (1964). Problems of optimization in learning a list of simple items. In M.W. Shelley & G.L. Bryan (Eds.), Human judgments and optimality. New York: John Wiley and Sons.

Suppes, P. (1967). Some theoretical models for mathematics learning. Journal of Research and Development in Education, 1, 5-22.

Suppes, P., Fletcher, J.D., & Zanotti, N. (1975). Performance models of American Indian students on computer-assisted instruction in elementary mathematics. Instructional Science, 4, 303-313.

Suppes, P., Fletcher, J.D., & Zanotti, N. (1976). Models of individual trajectories in computer-assisted instruction for deaf students. Journal of Educational Psychology, 68, 117-127.

Williams, M., Hollan, J., & Stevens, A. (1981). An overview of STEAMER: An advanced computer-assisted instruction system for propulsi[illegible]ring. Behavior Research Methods and Instrumentation, 13, 85-90.

The Present and Future Relationship of Technology and Testing

Frank B. Baker
Department of Educational Psychology
University of Wisconsin

I. Introduction

One of the hallmarks of the testing movement has been its long-term involvement with technology. In the early educational testing movement literature, the multiple choice item was referred to as the objective type item as its scoring involved no judgmental processes. Because of this, it was natural that efforts be made to mechanize the test scoring process. Over the years, many schemes were devised to efficiently score multiple choice tests via manual or mechanical devices of one type or another. While these devices worked, they could not really cope with the volume of test results being generated by the national testing programs established during the 1930's.

The first real technological breakthrough in the automation of test scoring was due to a high school physics teacher who invented a current summing device that could obtain a test score from pencil marks on an answer sheet. His invention became the IBM 805 test scoring machine of 1935 that was in widespread use until the 1960's. I am sure many of the "old timers" in the audience could regale us with a litany of the idiosyncracies of this machine.

However, as testing programs grew in size, the need for faster, more accurate automation grew. The second major technological advance occurred in 1955, and was the development of the optical mark reader

(Baker, 1971). This device could sense marks on the answer sheet optically and used the recently developed computer technology to score the test. Much of this scanning technology was developed by E.F. Lindquist at the University of Iowa, and resulted in his 1955 MRC machine which could process 6,000 answer sheets an hour.

Since that time, there has been a steady improvement in the accuracy, flexibility, and versatility of the OMR test scoring equipment. Today, there are a number of manufacturers of such equipment (Baker, 1983). Because of the capabilities and availability of such test scoring equipment, a wide variety of organizations (commercial testing companies, state education agencies, universities, and school districts) can conduct large scale testing programs in a cost-effective manner. As a result, one can consider the automated scoring and reporting of test results to be a rather mature field. However, it has provided the basic foundation for other applications of technology in testing.

II. State of the Art

The high capacity test scoring equipment represents one end of a range of test scoring equipment. At the other is the desk top scanner. In about 1967, Richard Schutz of SWRL explored the possibility of having a desk top scanner developed. However, events conspired to thwart this effort. It wasn't until 1974 that a commercially available desk top scanner (the DATUM 5098 OMR) was available. It could handle a 64 x 13 array of marking positions on an 8 1/2 x 11 sheet of paper, and cost $3,000.

This scanner was important as it could be connected in series

with a computer terminal and item responses could be sent over telephone lines to a central computer for scoring and reporting. The desk top scanners opened the door to cost-effective, small scale, automated test scoring and reporting. At the University of Wisconsin, we have DATUM scanner serial number 3, and it is still serving us faithfully for classroom testing. The DATUM 5098 was a very basic OMR and did little more than sense marks on the sheet and send item responses to a computer with an overall processing rate of about 15 sheets per minute.

In recent years, the microprocessor chip has become readily available, and desk top scanners have begun to incorporate them. The Europeans have been a few years ahead of us in this area and small scanners such as Kajaani Evalmatic use a microprocessor to compute and display the test score as the sheet is scanned. Recently, desk top scanners incorporating microprocessors have appeared in this country as well. Because these microprocessors are small but fast computers, they can be used to perform a number of functions within the scanner, one of which is the quality control of the mark sensing process. The microprocessor can obtain the readings from multiple marks, apply decision logic, and ascertain which mark is to be considered the student's item reponse. The microprocessor can score the test and in some scanners cause the score to be printed on the answer sheet. It is also possible to compute summary statistics as the sheets are processed. The microprocessors are also used to control the communications process and coordinate data with a computer. The inclusion of the microprocessor provides the desk top scanner with features and capabilities previously found only in the expensive high

capacity scoring systems. Yet, the overall cost of the desk top scanner has been quite stable for a number of years.

The linking of the desk top scanner to a microcomputer results in a small scale test storing and reporting system. Microcomputers such as the APPLE II, IBM PC, or TRS-80 have more than enough computing power to score a test, compute summary statistics, perform item analyses, and print a variety of reports. Thus, for about $5,000 a school can have an automated test scoring system that can process a very respectable number of tests per day. Recently, National Computer Systems has begun to market just such a system consisting of an IBM PC and an NCS 3000 desk top scanner. The Nippon Electric Company has been marketing a microcomputer-based test scoring system for a number of years. The system includes a mark sense card reader, a microcomputer, a typewriter printer, and an S-P keyboard. The latter is a special device used to enter item reponse data manually for a specific type of item analysis. Systems of this type should find widespread acceptance in the schools.

All of the test processing systems described to this point have used answer sheets upon which the students mark their item response choices. The Test Input Device (TID) (Syscon, 1983) is designed to eliminate the answer sheet and the OMR from the test processing sequence. The TID is about the size of an electronic calculator and has a similar keyboard and display. Internally it contains a microprocessor and about 20,000 bytes of computer memory. A probe on the device allows it to be connected to a computer via a "black box". Initially, the computer is used to prestore in the TID information about the test such as the test identification and the number of

items. In use, the student employs the keyboard to enter his ID number, the test ID, and then his response choice for each item. A feature of the device allows the student to review his choices, change them, and otherwise edit them via the keyboard and the display. When the student is done, he simply inserts the probe into the "black box", presses a button, and all the data are transmitted to the computer, where the usual programs score the test and report the results. The Navy is currently using the TID and appears to be pleased with its use as a test answering vehicle. The technology contained in the TID is a spinoff of military data collection, and the device itself is similar to those used to audit inventories in, say, a grocery store. Given the high cost of answer sheets, the TID could be quite economical to use in many testing situations.

A significant facet of any large scale testing program is the creation and maintenance of the item pool. This has led to the use of computers in this process, and a field known as item banking (Wood & Skurnik, 1969; Lippey, 1974) has arisen. The basic idea is to store the actual test items in mass storage and provide functions that allow one to inspect items, edit them in various ways, and select them for inclusion in an instrument. Since the reason that an item pool exists is to support the test construction process, most computer-based systems merge the item banking and test construction process into a single software package (Lippey, 1974).

A decade ago, I developed an item banking/test construction program based upon an item response theory approach that still defines the state of the art (Baker, 1972). This program maintained an item pool, kept historical records of the item and the test statisics, and

was integrated with a test scoring program. When a test was scored, all the historical data such as item and test statistics were automatically updated in the files. This computer program was implemented on a large scale computer and designed to be used by a relatively sophisticated test constructor. As one would expect, the item banking/test construction process has been implemented on a microcomputer (ATA, 1981). This system employs a TRS-80 with two floppy disks, and the computer program is called TEST BANK. The program can store about 300 items and the user can select items using a limited number of keys. Despite its modest capacity, this system represents a significant contribution to testing technology. As the data storage capabilities of the microcomputer grow, so will the size of the item banks that they can maintain. More importantly, such systems make item banking/test construction accessible to a very wide range of users.

Computer administered tests have been employed for many years, primarily via time sharing terminals from a large computer. Such testing has always been expensive and limited by the small number of terminals available on any one computer. The microcomputer has opened up new and interesting possibilities for on-line testing. It is quite possible to store test items on a floppy disk and have the microcomputer administer the test and record the student item reponses on a disk. Upon completion of a testing session, a student could be given immediate feedback by the computer as to the results and their interpretation in either a norm referenced or criterion referenced sense. In addition, the graphic capability of the micro can be used for figures and diagrams appearing within individual items. Voice

output devices such as the "type and talk" (Votrax, 1983) can be used to present questions and instructions verbally. Clearly, the microcomputer capabilities such as graphics, verbal output, data storage, and man-machine interaction can be forged into a very powerful and dynamic testing vehicle. The present state of the art offers all of these capabilities at a modest hardware cost. Microcomputer systems for administering standard psychological tests such as the MMPI are currently used in a number of settings. Such testing is also offered as a commercial service by a number of consulting psychologists.

One form of computer administered test is the tailored or adaptive test first proposed by Lord (1970) and investigated extensively by Weiss (1975). Under the adaptive testing procedure, a series of items are administered via a set of rules that select items that are appropriate to the examinee's ability. The available schemes narrow in on a student's ability level rather quickly, and each student gets a unique test. Such adaptive testing rests upon item response theory and is an excellent example of the application of the theory to practical testing. Given a precalibrated item pool, the actual item selection, test administration, and scoring procedures of the adaptive testing can be implemented easily upon a microcomputer. Such adaptive testing is particularly appropriate for schools using individualized instruction where students are tested on an as-needed basis rather than in groups at predetermined times. It also does not require personnel to administer the test. These and other advantages have led to the creation of microcomputer systems devoted specifically to adaptive testing. Vale and his associates in Minnesota are

currently developing such a system for the military.

Computer aided instruction and its derivatives, such as computer based instruction, have had a long and rather tortuous history. In the mid 1960's it was treated as the savior of education and considerable resources were devoted to the area without a great deal of success. By the mid 1970's the CAI furor had subsided and only a minimal interest existed. However, the emergence of the microcomputer, and its widespread penetration into the home and the schools, has given CAI/CBI a dramatic new lease on life, and it has risen Phoenix-like from the ashes. Numerous sources report that instructionally-oriented software for microcomputers is one of the major growth industries in the 1980's. Many new software houses have been established to produce such software, and old-line publishing houses are getting deeply invoved in the selling of instructional computer programs.

Inspection of this new generation of instructional software quickly reveals that it is nearly always devoid of any testing or evaluation component. Drill and practice programs will tell the student how many problems they got correct, but most other instructional computer programs do not. The general approach seems to be one of letting the student interact with the computer in various ways; when the student reaches the end of the program paradigm, the computer says goodbye and the student walks away. In addition, there are no records kept for the teacher to use in evaluating student progress. While such software may in fact be instructive, it does not integrate well into the overall educational process. The lack of a testing or evaluation component in such software is particularly

disconcerting.

At this point, I would like to briefly summarize the state of the art in testing technology before proceeding.

1. The desk top scanner and microcomputer combination provides the technological base for a cost-effective test scoring and reporting system. As a result, most organizations and schools can now afford the test processing associated with small to moderate scale testing programs. Such a capability allows schools much greater flexibility and control over their testing programs. The state of the art in high capacity test scoring and reporting is at a sophisticated level. Because of this, the data processing aspects of nationwide or other large scale testing programs can be performed efficiently and at a reasonable cost per student.
2. Due to microcomputers, item banking at the local level has become feasible. However, the practical size of the item pool is still somewhat limited by the available storage devices.
3. The existence of a dedicated hardware/software system for delivery and administration of tests is an important advance. In particular, the development of microcomputer based adaptive testing is an important increment to the state of the art in testing.
4. All indications suggest that the computer as an instructional device is finally going to make sufficient penetration into local schools to be considered a viable technology. However, existing instructional software rarely includes any provision for testing or evaluation.

III. Future Directions

A. Test theory trends

For those of us who work in psychometric theory, it has been clear for many years that classical test theory has run its course. This theory grew out of the work of Spearman on intelligence during the 1920's, and was the theoretical underpinning of the past half century of testing practice. However, a close look at classical test theory shows that other than some work on generalizability (summarized by Brennan, 1983), there has been little if any extension or elaboration of this theory in recent years. In addition, existing testing procedures and practices have fully exploited the capabilities of the theory. The theory is a mature one and its future growth does not seem likely. In sharp contrast, item response theory (IRT) is alive, dynamic, and growing rapidly. A considerable body of new theory is in place due to the efforts of Lawley (1943), Lord (1952), Bock (1972), Samejima (1972), Wright (1967), Wright and Stone (1979) and others. This theory is vastly superior to classical test theory in its conceptualization, as it is based upon the item rather than a test score, and has considerable potential for further growth.

At the present time, IRT is rapidly making the transition from a pure theory to one that is widely used in practice (Lord, 1970). In addition, it has provided analysis techniques for estimating parameters of items having graded or nominal response that classical theory could not handle. It is this author's view that the future will see an acceleration of the transition from a testing practice based upon classical test theory to one based upon item response theory. Because of its greater mathematical sophistication, the

application of item response theory is going to depend heavily upon technology.

B. Technological trends

The microcomputer revolution is not even a decade old, and the rate of change still appears to be accelerating. The early microcomputers were 8 bit machines having 8 bit address and data buses. Next we had the 16 bit machine with an 8 bit data bus. Today we have 16 bit machines with 32 bit internal registers and 16 bit address and data buses. The net result is that today's microcomputers are rapidly exceeding the capabilities of the medium scale computers of a few years ago, at a fraction of the hardware cost.

Although the hardware currently available has considerable power, that which is just around the corner is even more startling. For example, computers based upon the Intel 432 chip are just now beginning to emerge from the R and D shops. This family of computer chips allows one to create 32 bit microcomputers that are as powerful as today's large scale computers. Yet they are physically no bigger than the familiar personal computers. In a very real sense, this will be like having your university's computer center sitting on your desk. Thus, an increasingly large amount of raw computer power is becoming available to use within the context of testing programs and related applications. The only factor to dampen our enthusiasm for this class of machines is that the software to make use of it will cost many times that of the actual hardware. One may pay $15,000 for the computer and then $50,000 for the operating system, language compilers, mathematics libraries, statistical packages, etc., that are necessary to exploit the hardware.

The second area of very rapid technological advance is that of mass storage. The basic trend is toward ever larger storage capacity at a relatively decreasing cost per unit of storage. For example, most microcomputers have used a "floppy disk" that can store about 144,000 characters on a disk. With today's technology, this unit can be replaced with a unit of the same physical size that stores several million characters. Slightly larger yet modestly priced units can store from 10 to 40 million characters per disk.

It should be noted that a major limiting factor in a number of areas of testing such as item banking, test construction, and adaptive testing has been the lack of low cost mass storage. The increasing availability of low cost mass storage promises to have a significant impact on testing procedures dependent upon an item pool. Given the availability of such mass storage, the technical limitations on our ability to create and maintain such item pools are rapidly disappearing. It will also increase the feasibility of on-line adaptive testing as an item pool, for longer sequences of items can be accessible to the algorithms that select items for administration. From a practical application in testing point of view, the increase in low cost mass storage will probably have a greater impact upon the field than the increase in raw computing power. This is because most of our testing applications are heavily data based oriented and only marginally number crunching oriented.

Perhaps the ultimate in mass storage are the new optical storage devices. Some of these are "write once, then read only" devices. Once the data are stored they can never be changed. Such devices have considerable promise for archival data, such as obsolete item pools,

but could not be used for active item pools. However, optical mass storage devices that can be used like ordinary disk storage are beginning to appear and offer exciting possibilities in very large scale data storage. It will be a few years before this technology gets within the price range most testing organizations can afford.

A related memory device is the video disk which allows one to store full video screens and play them back. This opens up the door to dynamic presentation of test items such as the re-enactment of a historic event or the recording of a physical process. At the present time, the cost of material development and creation of a master disk is very high. However, copies of the disk are not too expensive. Thus, it would be possible to use this technology for testing, given the proper equipment. A few systems combining a microcomputer, a video disk player, and a color TV set are available commercially. But I am unaware of their use within the context of testing.

A third technological trend of interest is that in optical mark reading equipment. The introduction of the desk top scanner in 1974 has provided diverse organizations with a reasonably priced means of scanning answer sheets. There has been a trend toward greater sophistication within such scanners, but the cost has remained near the $3,000 per unit level. What is really needed is a low cost, say $600, desk top optical mark reader. Such a device is not as complicated as a printer, and the state of the art of optical marker reading is adequate; hence, to create and produce such a piece of equipment is not a major problem. However, I suspect that when such a scanner appears it will carry a Japanese nameplate. The availability of a really low cost scanner will put automated scoring and reporting

systems at the classroom level.

An age-old problem in the production of tests is including pictures and diagrams within the test item. Traditionally this has been done manually via cut and paste procedures. Fortunately, there are a wide range of commercial and military applications that face the same problem. For example, technical manuals are widely used and include both text and engineering diagrams that must be revised frequently to keep the manual current. To meet this need, equipment has been developed that can scan graphics material, convert it into a computer representation, and save it on a mass storage device. It is then a simple matter to merge these digital representations of the graphics with textual material and produce both the graphics and text on the screen of a video display terminal, or print them on paper.

It should be noted that a variation of this can be done with a personal computer using a GRAPPLER II board and an EPSON MX-80 printer with GRAFLEX chips. One can program the desired diagrams using high resolution graphics, and store the binary file on the floppy disk. The textual part of the question can be programmed and stored. To reproduce a hard copy of the item, the graphics information is read into memory and displayed on the VDT screen. With a single command, the diagram is reproduced by the printer. Then the text portion can be printed. This is not as nice as simply scanning the diagram, but it does provide a significant test item creation capability. Numerous computer programs, such as the Graphics Magician (Pelczarski, Lubar, & Jochumson, 1982) are available to facilitate the creation of the graphics part of the item.

One of the really active areas of technological development is

that of computer networks. The goal here is to create a communication network through which many different computers "talk" to each other. Where all the computers are in a reasonable proximity to each other, these schemes are called "local area networks". The driving force behind such networks is the automation of the office and the need for large corporations and/or government agencies to exchange data on a computer-to-computer basis. Perhaps the best known of the nationwide networks is the ARPA net that interconnects many universities and governmental laboratories. A major thrust is to interconnect microcomputers, and this can be done using commercial networks such as APPLENET.

Computer networks have a direct application to a wide range of testing procedures. For example, in the Netherlands an effort is underway to create a nationwide test processing network. Each school will be equipped with a desk top scanner, a microcomputer with disk drives, and a printer. When a testing program is conducted, the answer sheets are scanned locally, and the results stored on disks and then transmitted to a central computer. Upon scoring, item analysis, norming, etc., the test reports for the school and the individual students are to be transmitted back to the schools for printing via the microcomputer. At the same time, test results aggregated by school districts and other larger units are available at the central computer. While similar systems have been built in the past via the time-sharing capabilities of large computers, the microcomputer-based network offers much greater flexibility and ease of use.

Another application of computer networks is in on-line testing. It is now technically feasible for a central computer with some mass

storage to store many different item pools. A student can sign on to a microcomputer in the network, request to take a test, and have his request validated. The central computer will then transmit a test (possibly unique to the student) to the microcomputer, which then administers the items, scores the test, and tells the student the results. Upon completion, the item responses and other data are transmitted back to the central computer for aggregation and storage. Again, such a scheme has been possible in the past, but the microcomputer and local area networks place this within the realm of the readily achievable.

The final technological trend of interest to testing practice is in the area of software. Until quite recently, most applications implemented on a computer existed as separate computer programs with their own set of procedures and purposes. The disadvantage of this approach was that the user had to learn the procedures for each application in total isolation from all the other applications. Each had its own set of control functions, unique features that were tailored to the problem, and no commonality of logic. The result was a rather large learning period for a person who needed to use several different applications. About ten years ago, the trend was to place a common data base underneath these applications. Even though each application proceeded independently, they nonetheless used the common set of data. Within the past 2-3 years, efforts were initiated to integrate a number of seemingly disparate applications into a single coherent package. The first available integrated package was implemented on the APPLE LISA computer system. The software provides an integration of word processing, spread sheet calculations, business

graphics, list maintenance, PERT charting and computer terminal emulation. All of the data employed by a user of the system can be passed easily from one application to another via a simple user-friendly set of procedures.

A number of other such integrated systems are following on LISA's heels, and they will be commonplace in the near future. Although these integrated systems have been designed for the business environment, the basic approach and techniques are directly applicable to the testing environment. It is currently feasible to create a hardware/software system for a microcomputer that could integrate item writing, item banking, test construction, on-line administration of tests, test scoring, item analysis, and reporting of test results. Once the appropriate software tools are made available, this might even be done within the context of an existing system such as LISA. In any event, it could be done if one were to devote some people and resources to the task.

C. Instructional trends

The widespread penetration of the microcomputer into the schools is beginning to result in change in instructional approaches. At the college level, a large number of textbooks are being accompanied by a floppy disk which contains computer programs to be used in conjunction with the text. These progams range from simple exercises to sophisticated simulations of complex processes. In the physical sciences, many of these programs enable students to explore topics that would be prohibitively expensive to implement in a laboratory setting. A similar pattern is beginning to develop at the secondary and elementary level, where computer software is being used to

supplement existing texts and to provide enrichment. As mentioned in an earlier section, textbook publishers are moving quite rapidly to establish a market share in what is being called "electronic publishing". This activity suggests that the publishers see an underlying trend in which they must participate to ensure future business.

At the present time, the coordination between the textbooks and the computer software varies from very loose to a reasonably good level. Much of the software has been collected from a variety of sources and has been pooled under a commmon title rather than having been created specifically for the text. However, the longer term trend is toward a closer linkage between the material and approaches taken in the text and those in the computer software. At some point in time, instructional designers, curriculum specialists, psychometricians, textbook writers, and software developers are going to work as a team to jointly develop curriculum, instructional software, evaluation procedures, and instructional management systems, all within a common frame of reference.

It should be noted that the result will not be classical computer-aided instruction or computer-based instruction. Rather, the result will be textbooks written to take advantage of the educational leverage the computer can provide. Under this approach, an instructional decision is made that a specific topic can be handled better via the computer than by the textbooks or some other vehicle. Only then would a computer program be written and its use integrated into the instructional flow of the text. In many places, it would be clear that the text or other materials would be more appropriate. The

net result is going to be a mixed bag of conventional and computer-based procedures, all of which contribute to the instructional process defined by the text. In addition, the computer will maintain records of information needed by the teacher to effectively monitor and manage student progress through both the conventional and computer-based parts of the course.

The issue of present concern is where testing fits into such a highly integrated system. Conventional testing will be used to measure student progress in the broader sense. However, within the context of the computer-based aspects of the curriculum, a significant change will occur. The old "pretest, instruct, posttest, remediate" paradigm employed so widely in computer-aided instruction will be abandoned. This paradigm has been with us since the days of programmed instruction (Coulson, 1961), and is badly timeworn. The basic problem with the paradigm within the context of computer usage is that the student spends too much time answering multiple choice questions rather than using the computer in an optimal fashion. Present-day microcomputers have sufficient power to enable the student to use the computer as an exploratory tool as well as to implement highly dynamic modes of instruction. Time spent responding to multiple-choice items detracts from the student's productive use of the computer.

The testing alternative that I see developing is what could be called "non-intrusive" testing. Under this approach, a student using an instructional software package would never be administered a formal test within the context of using the computer. Instead of recognizable tests and test items, instructionally relevant data would

be collected as the student interacts dynamically with the instructional software. A variety of information, such as the sequence in which capabilities of the software are employed by the student, the rate at which a student moves from one level to the next of Bloom's taxonomy, and the strategies employed by the student to reach instructional goals, can be collected during the computer session. Given this data, evaluation routines embedded within the instructional software can ascertain the student's instructional status. An excellent example of this type of evaluation is the model developed by Brown and Burton (1978) to identify "bugs" in a student's learning procedures. When the desired level of understanding has been reached, the computer simply tells the student he knows the material and should move on. From the student's point of view, a test was never taken; however, from an evaluative point of view, the student has been continuously evaluated, and teachers have at their disposal a wide range of evaluation data collected by the computer.

Such non-intrusive testing clearly makes better use of the student's time as well as the computer resources. While many examples of non-intrusive testing procedures do not exist at present, there are some antecedents. The standardized grade score used by Suppes and Morningstar (1972) was computed dynamically as a student did drill and practice exercises. After each problem, this score was recomputed and used to select the next problem to be used. The diagnosis and prescription procedures implemented under CMI have similar characteristics even though they are based primarily on test scores (Baker, 1978). The analysis portions of non-intrusive testing have much in common with the current efforts in artificial intelligence and

and regardless of the level of testing, the technology for processing the results can handle the workload.

Modern technology has also provided the means for maintaining very large item pools, and providing automated or semi-automated item selection from these pools. It is not uncommon to have pools of up to 25,000 items stored via a computer system. The quality of these item pools is yet another matter. Whenever such a large pool exists, it is usually the result of having many people in many different settings write items and enter them into the pool. In such circumstances, it is extremely difficult to maintain the instructional focus of the items as well as technical quality. For example, Brenner (1981) reported, in the case of a pharmacology item pool of 25,000 items, that only 6,500 items were retained after scrutiny by a review panel of subject matter specialists. Most large item pools would exhibit similar shrinkage upon close inspection. The underlying message here is that technology can make a process easy to implement, but it does not ensure the quality of the product.

Technology has contributed indirectly to a new testing problem. In an earlier era, the development of tests was a rather academic process. Typically, a scholar became interested in a subject matter area and constructed a test. The instrument was refined in the course of a few school years. Graduate students would further explore the instrument through their thesis research. After a few years, the instrument was reasonably well developed, and some research existed that described the reliability, validity, and interpretation of the instrument. In some cases, the scholar printed and distributed the test in the marketplace. In other cases, the test would be taken over

in particular with "expert systems". The latter are computer hardware/software systems that embed the decision-making heuristics of experts within the software, which is then used by less skilled persons to analyze situations and draw conclusions. The immediate example is medical diagnosis, but educational diagnosis could just as well have been the area of interest implemented via an expert system.

D. Summary of trends

Again, let me briefly summarize the future directions as I see them.

1. Clearly, classical test theory has reached its upper limit of development and will be replaced in practice by item response theory.
2. Microprocessor-based technology will continue to move at a rapid pace. This will make it easier to automate existing practice. It will also provide the basis for developing specialized hardware/software systems for use in the field of testing.
3. There will be an increased emphasis upon a coordinated approach to instructional design that exploits the microcomputer as an educational vehicle. Non-intrusive testing could be a significant part of this approach.

IV. Barriers and Problems

A. Conventional testing

Let me mention several old problems before describing a new one. Modern technology, in the form of desk top scanners, microcomputers and high-capacity test scoring equipment, makes the automation of test scoring and reporting easy and cost-effective. However, there are limits to the amount of testing that the schools can absorb. We have seen several swings of the testing pendulum in the past few decades,

by a commercial testing organization and marketed. Even when instruments are developed within a testing company, the basic paradigm holds with a somewhat different cast of workers. On the whole, test development was a rather gentlemanly pursuit that was only sporadically impinged upon by forces outside of the educational establishment.

Beginning about 20 years ago, in the context of employment screening instruments, political forces have intruded into testing. At the present time, two politically motivated events have occurred that have a major impact upon testing. First, the competency testing of the 1920's and 30's has been resurrected from the grave and given new life. Second, in some states, item disclosure laws have been passed which give examinees access to items in the tests they have taken. I don't intend to argue the merits of these two, but the process by which they came about and their impact upon testing bear some examination. The politicians who pushed these two ideas did so with little or no understanding of the nature of test development, the technical issues, or the long-term ramifications for educational measurement. The net result is that measurement specialists have been thrust into a situation for which they are ill prepared.

In the case of competency testing, there is a demand for immediate large-scale testing on a wide range of subject matter, with little or no underlying test development. The item disclosure rules result in tests that have been carefully developed over many years being put in the public domain. If the test is to be retained, it forces those reponsible for the test into a high-speed, iterative item development process to replace compromised items. Although there are

many forces contributing to an increased role of the politicians in the measurement arena, one of the culprits is technology. The general public's naivety about how computers work carries over into the political arena. What happens is that the politicians <u>hear</u> that tests are scored by computers, that item pools are maintained on the computer, and that tests can be printed via computer. The conclusion reached is that if so much of testing is automated, it must be a simple matter to establish a testing program; you just let the computer do it! I suspect that in the long term, the ability of measurement specialists to cope with these external inputs will depend upon that same technology that helped get them in trouble.

B. Technological barriers and problems

Technology itself is also a barrier and a problem. Because of the rapid pace of technological development, the hardware in particular advances faster than we are able to absorb it into the daily world of testing. Taking advantage of new technology, such as optical storage, involves major levels of effort. It takes time and money to explore what testing uses can be made of the technology, and the start-up costs are independent of the eventual use of the technology in the schools. In addition, making the transition from a feasible use of technology to one that is widely used in the schools involves a very high level of effort and cost. Even when such technology does reach the schools, it requires resources which are scarce. Hardware must be maintained. It becomes technologically obsolescent rather quickly, and the trade-in value of an old piece of hardware is nil. Thus schools have to seriously consider the cost-benefits ratio when introducing technology of any kind. In

particular, any testing-related technology must have a favorable cost-benefit ratio in order for it to be widely adopted by the schools.

While the creation of hardware/software systems that integrate item writing, item banking, test construction, on-line or adaptive testing, graphical capabilities, automated test scoring, and reporting will occur, it is a significant development task. For example, the LISA system is reported to have cost $50 million to develop. An integrated testing system is at an equal level of technical complexity, yet the potential market for such a system is very small compared to that for a work station such as LISA. As a result, the development of such systems is going to be an evolutionary process based upon available technology rather than a sudden quantum step. The components are all there; it is their integration into a flexible, powerful system with sufficient generality that costs time and money.

C. The manpower barrier

The current testing milieu is one in which testing is not as static as it once was. The field is more dynamic, more dependent upon technology, and increasingly under greater scrutiny. More importantly, the context within which testing is employed is becoming increasingly unstable. Because of this, it is increasingly difficult to construct items and to refine and polish them in a volume necessary to meet the need. One consequence of this will be a lowering of the quality of the tests due to insufficient development. The conclusion is that maintenance of quality both in terms of content and technical characteristics requires more trained personnel than are currently devoted to test development. In addition, if testing is going to

exploit technology, persons are needed who can work within both measurement and hardware/software. This is particularly crucial if adequate integrated sysems are to be developed to support testing. The trained manpower barrier to future development of testing is not an obvious one, due to the diffusion of such manpower across a wide spectrum of levels in government and educational institutions. Yet it does exist.

V. Implications and Policy Recommendations

A. Education

1. The rapid transition from classical test theory to IRT has many implications for the use of tests in the schools (specifically, the reporting of test results in an ability metric that will do much to improve the interpretation of test results). IRT allows the use of new item types as well as new testing procedures such as adaptive testing. The policy recommendation is twofold. First, schools and other responsible agencies should foster the switch in the psychometric underpinnings of testing practice. Second, vehicles need to be put in place to fully explore the ramifications of item response theory for the day-to-day pratice of testing.

2. The increasing use of microcomputers in the classroom opens the door to the use of "non-intrusive" testing procedures. However, this idea is not well formed at present, and considerable research needs to be done to determine the underlying principles of such testing. Without such research, each application of the approach is a special case, and it will be difficult to determine if the basic concept is viable. Once the basic framework of non-intrusive testing is established, the incorporation of such testing into modern

computer-supported curriculum will be much easier.

3. Test development

Conventional testing as we now know it is going to be with the schools for a long time. However, the demands upon the schools for both externally and internally imposed testing will increase with time. The major implication is that there will in all probability be a continuing shortage of personnel trained in measurement and its related technology to develop such tests.

One of the clear outcomes of the CSE Study of Test Use is that the majority of classroom testing employs teacher-made tests. Yet, classroom teachers are provided with very little assistance in the preparation of such tests. Thus, efforts should be initiated to develop a microcomputer-based test development system for use by classroom teachers. Such a system is within the state of the art, and its availability could have a significant impact upon testing in the schools.

B. Technology

There currently is a significant lag between the introduction of a new level of technology and its application to the field of testing. What is needed are vehicles so that this technology can be employed quickly and its advantages/disadvantages for use in testing can be determined. Such early evaluation allows one to both discover viable uses of such technology and to enable others to avoid nonproductive uses of the technology. Let me briefly describe some examples of areas of technology where pilot projects would be valuable.

1. The availability of low-cost mass storage for microcomputers

has major implications for item writing, item banking, test construction, and on-line adaptive testing as well as for automation of test scoring and reporting. We need to look into what can be accomplished using these storage devices. In particular, their role in the development of tests needs to be explored.

2. Video disk technology opens many possibilities in both conventional and non-conventional testing procedures. It makes possible test items that involve dynamic presentations as well as active examinee participation in the evaluation process. The Achilles heel of this technology is the enormous material preparation time. The role of video disk technology in testing, as well as vehicles for minimizing the material preparation time, needs to be examined.

3. The leading edge of the application of computer technology currently deals with computer networks. The hardware and software is available to construct a wide variety of networks. Such networks have many implications for testing. These involve down loading of tests to local sites from central sites, aggregation of test results across widely distrbuted sites, and flexible mixes of conventional on-line and adaptive testing. The interesting feature of this work is that it is focused upon allowing microcomputers to be networked. The ramifications of networking for testing need to be investigated.

Pilot projects in these and other areas can be conducted in a variety of settings and are within the capabilities of a range of educational institutions. The results from such pilot projects would do much to set the tone for the improvement of testing via technology.

VI. Summary

The intent of the present paper was to provide an overview of the

symbiotic relationship between testing and technology. This relationship has been developing since the earliest days of the testing movement. Despite the age of this relationship, it has not gone awry. One of the major factors in the continuity of this relationship is that the cost of high technology has been reduced to the point where it is accessible to most of those with an interest in testing. Because of this, one is as likely to see sophisticated research and development projects dealing with testing at the local school district level as in professional educational innovation organizations. As a result, these are rather exciting times in the field of educational measurement. Hopefully, one outcome of this conference will be the addition of further excitement to the relationship of testing and technology.

References

Baker, F.B. (1971). Automation of test scanning, reporting and analysis. In R.L. Thorndike (Ed.), Educational Measurement, Washington DC: American Council in Education.

Baker, F.B. (1972). A conversational item banking and test construction system. Proceedings of the Fall Joint Computer Conference, AFIPS, 41, 661-667.

Baker, F.B. (1978). Computer managed instruction: Theory and practice. Englewood Cliffs: Educational Technology Publications.

Baker, F.B. (1983). Automated scoring systems. In T. Husen & T.N. Postlethwaite (Eds.), International encyclopedia of education: Research and studies. Oxford: Pergamon Press.

Bock, R.D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. Psykometrika, 37, 29-51.

Brennan, R.L. (1983). Elements of generalizability theory. Iowa City: American College Testing Publications.

Brenner, L.P. (1980-81). On-line item banking in health sciences education. Journal of Educational Technology Systems, 9, 213-227.

Brown, J.S., & Burton, R.R. (1978). Diagnostic models for procedural bugs in basic mathematical skills. Cognitive Science, 2, 71-109.

Coulsen, J.E. (Ed.). (1961). Programmed learning and computer based instruction. New York: Wiley.

Lawley, D.N. (1943). On problems connected with item selection and test construction. Proceedings of the Royal Society of Edinburgh, 61A, 273-287.

Lippey, G. (Ed.). (1974). Computer-assisted test construction. Englewood Cliffs ,NJ: Educational Testing Publications.

Lord, F.M. (1952). A theory of test scores. (Psychometric Monographs No. 7).

Lord, F.M. (1970). Some test theory for tailored testing. In R. Holtzman (Ed.), Computer-assisted instruction, testing, and guidance. New York: Harper and Row.

Lord, F.M. (1980). Applications of item response theory to practical testing problems. Hillsdale, NJ: Erlbaum.

Marco, G.L. (1980). Item characteristic curve solutions to three intractable testing problems. Journal of Educational Measurement, 14, 139-160.

Pelczarski, M., Lubar, D., & Jochumson, C. (1982). The graphics magician. Geneva, IL: Penguin Software.

Samejima, F. (1972). A general model for free response. [Monograph 18]. Psychometrika, 1972, 37, Part 2, 1-68.

Suppes, P.C. & Morningstar, M. (1972). Computer-assisted instruction at Stanford, 1966-1968: Data, models, and evaluation of the arithmetic programs. New York: Academic Press.

The Test Bank. (1981). San Diego: Advanced Technology Applications.

The Test Input Device. (1983). Washington, DC: Syscon Corp.

Type 'N Talk User's Manual. (1981). Vortrax.

Weiss, D.J. (Ed.).(1975). Computerized adaptive trait measurement: Problems and perspectives. (Research Report 75-5). University of Minnesota, Psychometric Methods Program, Department of Psychology.

Wood, R. & Shurnik, L.S. (1969). Item banking. London: National Foundation for Educational Research in England and Wales.

Wright, B.D. (1968). Sample-free test calibration and person measurement. In Proceedings of the 1967 Invitational Conference on Testing Problems (pp. 85-101). Princeton NJ: Educational Testing Service.

Wright, B.D., & Stone, M.H. (1979). Best test design. Chicago: MESA Press.

From Domain-Referenced Curriculum Evaluation
To Selection of Educational Microcomputer Software

Wells Hively

President, Hiveley's Choice Publications, Inc.

Much of my past work has been in the field of domain-referenced testing and curriculum evaluation. Some of that work took place in a happy association with the UCLA Center for the Study of Evaluation, which published one of our contributions to this field as the first of its Monograph Series in Evaluation (Hively, Maxwell, Rabehl, Sension, & Lundin, 1973). Those of you who know this work probably will not be surprised at the approach we are now taking to the selection of educational software: compare, contrast, classify, and try to avoid over-generalization.

Currently, we are concerned with evaluating microcomputer programs that can enhance instruction during the period of schooling when it is easiest to consider the curriculum as a whole: preschool through grade 9. We have formed a publishing company to assemble and transmit information about educational microcomputer software to schools. Our purpose is to help school people more easily find what they need and use it more effectively. Specifically, we want to help teachers answer the following questions:

What kinds of programs are currently being developed?

How can we find good ones?

How can we use them effectively in school?

How can we tie them into the basic school curriculum?

We assume that there are many different types of educationally useful microcomputer programs, each with its own practical purpose,

each derived from its own theoretical assumptions and each, therefore, requiring its own unique set of evaluative criteria. We also assume that the lesson-plan settings in which teachers use the programs have at least as much influence on their impact as the characteristics of the programs themselves. Consequently, useful evaluation must take account of both the characteristics of the programs and the ways in which they are used.

The terminology used to classify different types of programs has by no means settled down. To make matters worse, the terms often carry evaluative connotations. Currently, outside the military, "CAI" (computer-assisted instruction) is a low-status term. "Drill and practice" is out. "Simulations" are in. "Learning games" are in. "Computer literacy" is definitely in. But all these terms are operationally hazy. It's important to try to clarify the terms we use to classify programs, because the classifications govern our approach to evaluation: programs are relatively easy to compare and evaluate within the same class, but very difficult to compare across classes. Let me give you examples of the different classes of programs we are encountering, and suggest some more precise nomenclature for them.

Types of Programs

A simple and generally useful type of program may be called "Domains of Practice". A good example is a program published by Sunburst Communications called Smart Shopper Marathon. The purpose of the program is to provide practice in rapid arithmetical estimation. The setting is an imaginary supermarket. Each so-called "aisle" represents a different domain of practice. In aisle 1, the student

has to rapidly estimate the results of dividing a decimal by a whole number. The student's job is to answer as many problems as possible in a given time, so quickly that detailed calculation is a hindrance: rounding off and estimating is the skill that must be practiced. In aisle 2, the student has to estimate the results of multiplying whole numbers. In aisle 3, subtracting decimals; in aisle 4, comparing fractions; in aisle 5, multiplying whole numbers times decimals.

The problems in each set are generated in random order, and each time you use the program the "aisles" appear in a different order. Therefore, because students are not likely to memorize rote sequences, the program lends itself to repeated practice without boredom. The scores used to judge youngsters' progress are based on a combination of speed and accuracy.

Programs like Smart Shopper Marathon are characterized by items drawn from clearly defined domains of knowledge or skill, a high frequency of opportunities to respond per unit time, and almost total absence of instruction presented by the program itself. The teaching of the constituent skills and the orientation to the problem-solving approach must come from an outside source. The students must obtain guidance in strategies for estimation from their teacher, or from each other, working in a small group. Therefore, programs like this make good vehicles for classroom demonstrations and for small group discussion. They provide external focus and feedback around which classroom activities can be assembled. Increased performance on the domains of practice presented by the computer may become the criterion toward which teacher and students can work together, a welcome change from the teacher's usual job of standard setter and task master.

It is useful to compare programs of the domains-of-practice type with programs of a second type that have historically been called "tutorial". A tutorial program developed by the Minnesota Education Computing Consortium leads up to the geometrical definition of an angle. What the student encounters in programs of this class is a fixed and predetermined sequence of presentations of bits of information interspersed with questions and answers. The term "tutorial" is too broad to clearly denote this type of program. Let's narrow the terminology to "linear tutorial". This is the classic programmed-instruction format which most people associate with the so-called "CAI" that is currently out of fashion. Pehaps one can see the reason why. The frequency of opportunities for students to respond in linear tutorial programs is low in comparison to the rapid-fire opportunities provided in domains of practice. Nearly always, the expository material could be conveyed faster in a book or in a conversation with a teacher or a peer. Perhaps most important, the sequence of "telling and testing" arises totally out of the mind of the designer, with no elbow room given for the idiosyncracies of different learners.

In constructing domains of practice, we are on fairly safe ground because we are creating models of subject matter. The theory and methodology of domain-referenced testing provide a fairly solid foundation for this job. But in linear-tutorial programs we are attempting to model the dynamics of teacher-student interaction without actually <u>allowing</u> dynamic interaction. There is little theory to guide this task, and successful programs of this kind are hard to find. Perhaps artificial-intelligence theory will eventually help us

construct truly dynamic tutorial programs of the Socratic or error-analysis type. But good, dynamic tutorial programs have not yet filtered down to the practical level where we encounter them in our survey. For practical purposes, teachers can do much better by putting small groups of children to work on domains-of-practice programs, and letting the teaching arise from class discussion and spontaneous interaction, than by sending individual children off to have learning doled out to them in small droplets by step-by-step, linear-tutorial programs.

An enormously popular third type of program is the education game similar to those seen in video parlors and arcades. Basically, these games are domains-of-practice with several added attractions. This combination may be called extrinsically-motivated practice or extrinsic games. Some of them are a lot of fun.

An example of an extrinsic game is the DLM Company's Alligator Mix. If your answer in the alligator's belly matches the problem in the apple that comes floating in from the left side of the screen, you win by opening the alligator's mouth and swallowing it. If it doesn't, you leave the mouth closed, and it bounces off and spins away. At the beginning, there is just one alligator, at the bottom of the screen, and the apple has to travel a long way, so you have plenty of time to make up your mind. After a string of successes, a new alligator surfaces. The apple doesn't travel as far to get to that alligator, so you have to think faster. When there are four alligators lined up, you really have to hustle. The teacher or student can choose from nine skill levels, which have to do with the velocity of the apple's motion, and three problem ranges which have to

do with different domains of practice.

Another example from DLM is called Demolition Division. The tanks all come forward at the same time, shooting at your forts. Your job is to position a 0 underneath the cannon that aims at the most threatening tank. Then change the 0 to a number that corresponds to the questions on the tank, press the space bar, and destroy the tank before it knocks down the wall and destroys your cannon.

What makes these games fun is delicate grading of speeds and levels, freedom to select levels that match entering skill, and richness of alternatives in ways to respond. A whole art and technical literature is growing up in the area, and standard "plot formulas" are rapidly appearing.

Another kind of plot formula for an extrinsic game is demonstrated by Sunburst's Math Mansion. It gets good mileage out of a "dungeons and dragons" theme. The thematic development and the richness of alternatives in Math Mansion trade off against relatively low frequencies of opportunities to respond. We are a long way from knowing, if we ever will know, what are the optimum mixes of such ingredients. But youngsters identify the good examples by their attention and their resultant learning.

A fourth category of program might be called, by way of contrast, intrinsic games. QED Company's Arith-magic program is called "Diffy". The student volunteers a set of four numbers which the computer places at the corner of a square. Then the student goes around the square finding the differences between each pair of numbers. The differences found in the first round then form the

corners of a new square for the second round, and the student goes around finding the differences again. This goes on until, eventually, lo and behold, the differences all come out the same. The challenge is to figure out what characteristics of the starting numbers make the differences converge quickly or slowly. The game provides a vehicle for discussion, exploration, and curiosity, and incidentally provides a very high frequency of opportunities for subtraction practice.

Other examples of intrinsic games are Sunburst's Teasers by Tobbs and MECC's Taxman and Bagels. Games like these tap into the whole realm of classic puzzles and brain twisters, some of which lend themselves nicely to computer presentation. As usual, the most frequent examples tend to be in the field of mathematics, but there is no reason why they need to be limited to that field.

A fifth promising category of programs is exemplified by two in the Milliken Company's Edufun series: Golf and The Jar Game. Let's call them intuition-building programs. In Golf the problem is to direct the ball from the tee to the green by estimating an angle and a distance using a compass rose for reference, and a given unit of distance. If you lead off, you must estimate distance and direction absolutely (in terms of the compass rose and the unit of distance), but if you shoot second, you can correct your direction by adding or subtracting degrees to the course taken by your opponent's ball, and you can correct your opponent's estimate of distance. The game builds up a nice intuitive judgment of angles and directions.

In a similar vein, The Jar Game builds upon the intuitive statistical notion of drawing beads out of a jar. The young student is then shown diagrams of jars containing different proportions of two

kinds of candy pieces. The job is to figure out on which jar of candy a randomly-directed fly will land more often.

There are many other potential examples of geometrical and statistical intuition-building activities that computer experiences could enhance. The ease and speed with which the computer can generate these examples is delightful. We have come a long way from the old days of having children estimate the number of raisins in an average slice of raisin bread by taking apart a loaf of bread and counting the raisins in selected slices.

The sixth category is simulation programs. There are so many different kinds of simulations, and they can produce so many different outcomes, that this category will no doubt be subdivided later, but the characteristics that guide subdivision are not yet clear. The MECC <u>Sell Series</u>, built around the famous <u>Sell Lemonade</u>, is an example. The simplest one of the series is called <u>Sell Apples</u>.

When youngsters are turned loose on a program like this, they may learn many different things, depending on the context provided by the teacher. They <u>may</u> learn to read carefully and follow instructions in detail. They <u>may</u> learn to interpret data in tabular form. They <u>may</u> build up an intuition about the relationship beween price and volume of sales. They <u>may</u> learn important habits of record keeping. They <u>may</u> learn to transform data into graphical form and interpret trends. At a deep level, they <u>may</u> learn some important strategies that underlie scientific method, such as choosing extremes of variables and narrowing down to find the maxima and minima. They <u>may</u> even learn something about the cost benefit of seeking truth. None of these

things is taught for sure by the program. They depend on the context provided by the teacher and other students. It is particularly obvious, in the case of simulation programs, that validity and usefulness depend as much on the context provided by the teacher and peers as on the programs themselves.

MECC also provides a nice example of the seventh category: information retrieval programs. Nutrition asks you to provide a list of your food intake for one day. Then it gives you a nutritional analysis: how well your day's food intake represents the four basic food groups, how your numbers of calories provided by fat, carbohydrate, and protein compare to the recommended number of calories for a person of your age and stature, and how your intake of iron, calcium, vitamin A, and vitamin C compares to the recommended daily requirements. This is what the computer does best, and its role in this kind of instruction is distinctive. MECC's Nutrition program does not provide a means of accessing or adding to the nutritional data base, but one can easily imagine programs to which teachers and students might add information for foods not currently included, or ask for other kinds of analyses; a nice meeting ground between specific subject matter and computer literacy.

Nutrition is a miniature data base, and elsewhere a wonderful array of useful data bases is becoming available to teachers and students. Compuserve, for example, is a service that enables computer owners to obtain information from many data bases at night--airline schedules, weather rpeorts, and so on. The potential of data bases such as these as vehicles for instruction is tremendous. Answering

questions that come up in class by accessing a nutritional data base overnight would be a considerable step up from writing letters to the Department of Health.

The eighth and last category is such a large and heterogenous category that it, too, will no doubt soon be subdivided. For now, let us call it "tools and displays". In this category are all the programs that perform helpful calculations, the programs that process words, programs that display graphs of changes in phenomena detected by sensing devices like thermometers, and programs like the famous Logo that offer environments with important properties to be explored. The educational utility of these programs is limited only by imagination and experience. The following is just one example.

A program produced by Spinnaker Software called Delta Drawing is a kind of baby Logo. The commands are easy to understand, and a small child can start making interesting pictures almost immediately. We start with an arrow, move it forward by pressing the D key, and turn it to the right by pressing the R key. We change the color of the line by pressing the C key and then a number corresponding to the color we want. We move it forward again and change its direction and color. We store all the preceeeding steps in a sub-routine which can be repeated. We repeat the sub-routine to generate a kind of rose window. We may fill the spaces on the screen by choosing a color an then pressing Control F. The computer keeps track of all these steps as a string of symbols, and we can switch to text mode from the graphics mode and examine the string, operate on it, and go back to graphics to see the results.

Programs like Delta Drawing offer nice opportunities to explore symmetries and artistic effects. For example, when a line passes beyond the border of the screen, we have a choice of having it wrap around and appear from the opposite border or having it bounce off at an equal angle. The line bounces and bounces again like a billiard ball. It continues bouncing to generate a symmetrical pattern.

It is also possible for a repeated figure to wrap around and then bounce to produce a complicated effect. It goes on bouncing and creates an interesting artistic result made up of a combination of expected and surprising features. There is considerable potential in programs like these in the hands of teachers with sensitivity to some of the relationships between art and mathematics.

The overwhelming impression one gets from watching children work with all the foregoing different types of programs--ranging from open-ended environments, like Delta Drawing, to practice sequences like Smart Shopper Marathon--is that their effectiveness depends at least as much on the classroom context in which they are used as on the properties of the programs themselves. Properties of programs which are drawbacks in one context may be benefits in another, and exciting uses may be totally unanticipated by the people who developed the programs.

A Curriculum Guide

With the foregoing review of types of programs as background, now let me tell you about the product of our work: a curriculum guide for grades 0-9 called Hively's Choice. The target audience is what you might think of as "second wave" educators--not the original enthusiasts, but the experienced and thoughtful mainstream teachers on

whom any successful educational innovation depends.

Several characteristics distinguish the guide from other efforts to help teachers evaluate and choose software. First, the guide only contains software that has been found to be particularly outstanding in quality and ease of use. Second, the guide is designed in such a way as to make it as easy as possible for teachers to connect the recommended software to curricula and lesson plans. This is done in several ways. The user may begin by looking at a chart showing where each of the programs fits into general subject matter areas and the grade levels over which it is likely to be useful. Next, the user turns to a set of quick descriptions, organized by subject matter, within grade levels, and arranged so that one can look through them very rapidly so as to maximize chances of discovering unexpected connections to upcoming lesson plans.

The reader who finds something of interest by perusing the quick descriptions may turn to a detailed description of that program. There, the goal is to describe the program in enough detail that one can intelligently decide whether it would really be useful and exactly how it would relate to ongoing curriculum.

A subsection of the detailed description called "Curriculum Connections" includes words and phrases that can be used as cross references to scopes and sequences. "Objectives" briefly describes the kinds of learning which may be enhanced by the program, and a section called "Instructional Examples" gives recommendations about how best to utilize the program in classroom discussion, small group or individual work. "Estimated Engagement Time" helps teachers plan how much time to allow for work on the program by the whole class,

small groups, or individuals.

Also in the detailed description section one may find the technical information about the program, the hardware it requires, and the name, address and telephone number of its producer. The rest of the book consists of cross indices by subject matter and topic, and an alphabetical listing to facilitate the location of specific programs.

Our goal is to find rich, engaging and easily usable programs that have solid connections to all the areas of the basic curriculum, preschool through grade 9, and that take all the different forms described in the earlier part of this paper. If you imagine the curriculum as a matrix of subject matter areas by grade levels, some of the cells in the matrix are already getting crowded while others are virtually empty. Over time, our goal is to weed out programs in the crowded cells so as to include only a selection of the most useful and interesting ones, and to seek entries in the empty cells. Each year the guide book will be revised, following the analogy of a European travel guide. Like a travel guide, each edition will be cumulative and self-contained.

Organizationally, this work is done by a small, carefully selected group of contributing editors, who work in schools and work very closely with teachers in training. These editors are chosen to represent areas of the country where thoughtful and interesting work with microcomputers is going on. In their daily work with teachers, the contributing editors keep an eye out for outstanding programs and interesting ways of using them. They forward their

reviews to a small central editorial staff that produces the book.

In meetings of this editorial staff, we work to explicate the bases for our selections, to clarify categories of programs and the evaluative criteria applicable to each, to organize observations about effective classroom use of various types of programs, and to identify useful sequences and combinations of programs. This work aims to create a dialogue between good theory from the technical literature and careful observations of classroom use. From these, we are developing, year by year, a progressively more useful, readable, and balanced curriculum guide--one which can contribute both substance and diversity to the curriculum for preschool through the first nine grades.

Reference

Hively, W., Maxwell, G., Rabehl, G., Sension, D., & Lundin, S. (1973). Domain-referenced curriculum evaluation. Los Angeles: UCLA, Center for the Study of Evaluation.

Implications of New Technology for Social Policy

Robert M. Hayes

Dean, Graduate School of Library and Information Science

University of California, Los Angeles

I'm not sure that I can provide the kind of focus that was wanted for this meeting, since my own work has been totally in the context of university teaching and research concerning computer applications to information storage and retrieval, their effects upon libraries and the information industries, and related public policy. The problems and concerns faced by the broad range of educators, especially with respect to issues about testing and evaluation of programs and students, are therefore considerably outside my area of expertise.

However, there is a topic of special concern to me that may provide you with a springboard for your own discussion. It has been characterized as the "information age", and I think it ought to be of concern to you. Therefore, in my comments this afternoon, I plan to sketch out the broad outlines of the information age, and then to pose some questions about the responsibilities of all educators with respect to preparing students for their lives in this new world. The issues which are of paramount importance, the ones for which the coming generation must be prepared, are those in social policies of both government and the private sector.

The Information Age

The general context is provided by the report "The Information Economy" of the Office of Telecommunications. It presents data that show that 50% of the nation's workforce today is directly engaged in

information activities. Of them, 25% are in "information industries" and 25% are in the internal information operations of companies.

The crucial thing is not so much the magnitude (50%), since the definition used in that report is so broad that it's bound to include a high percentage, but the growth rate in that percentage over the past 100 years and more. The percentage engaged in information activities, taken in the broadest sense, has increased from less than 5% to 50% within the past ten decades--clear growth at rates of 25% per decade. In fact, the match between that exponential growth rate and the actual growth is frighteningly close. Furthermore, when one also recognizes that the working population has increased exponentially during that same period, the growth in number of persons in information work is awesome! Since prior to the report on "The Information Economy" this fact of exponential growth in information activities was buried in the data for the "services sector" of the economy, it is only now that policymakers are beginning to recognize the magnitude of this phenomenon and to be concerned about what it means in both social and economic conditions.

The report "The Information Economy" defined "information industry" in the broadest possible manner, encompassing not only traditional agencies for distribution of information (publishing, libraries, scientific and technical information centers), and all of the information technologies (computers and telecommunications), but all of banking and insurance, all of education, all of real estate and stock and bond brokerage, all of advertising, all of the entertainment industry. Of course, once the initial shock of such a broad definition has been felt, it quickly becomes clear that indeed these

all are "information" industries. In fact, with any rational definition of the term it is virtually impossible to limit the scope more narrowly.

It seems clear, in light of what is happening with the growth of information activities, that information is an economic entity. Costs are incurred in producing it; it can be used for a variety of purposes that have economic value, so people are willing to pay for it; it can serve as support to producing products and services, and thus behaves somewhat like a capital resource; when used by the information industry itself, it actually does serve as a capital resource, a means of production of products and services rather than simply an ancillary support or management tool.

These are the aspects that I want to emphasize in this talk, but before doing so, I must recognize the technological aspects: communications, computers, the people and facilities needed to maintain such equipments. In fact, most discussions of the nation's information systems are stimulated by, and start from, considerations of these technological issues. And rightly so, since they indeed are spectacular. Each of the information technologies has undergone dramatic decreases in cost. And there do not appear to be any fundamental barriers to continuation of those past rates of improvement. It is also clear that the continued growth of the information economy will depend upon and be fed by the information technologies. Indeed, the very title of this conference itself emphasizes the technological aspects as they relate to the topic of testing, the primary concern, I suspect, of most of those attending.

But these technologies--computing, data retrieval,

communications--are not, in fact, the primary issues. It is clear that societal forces are what have made information increasingly important. In particular, increasing size and complexity of organizations, growth in populations, and technological revolutions all combine to make improved information essential; otherwise, the organization could not be managed, the communication among people would become more and more difficult, and technical development would cease. It is therefore an imperative of life that there be a growth in information institutions as the means to meet those needs.

A key point, related to the technologies, is the fact that while the effects of the technology are to reduce the costs of the physical carrier of information (whether in the form of a printed publication, a digital data base, or a telecommunication), that doesn't mean that the costs of the information package itself will be comparably reduced. What it does mean is that the intellectual costs--the real "information" costs, the costs involved in creating the information itself, in selection and quality control, in marketing and distribution--will become an ever-increasing proportion of the total costs. Those are the real functions of the publisher (in contrast to the printer), of the data base service (in contrast to the computer), of the library (in contrast to the mere collection of books). This fact needs to be recognized if the technological effects are to be seen in proper perspective. And it means that the organizations involved in the information transfer process will become increasingly important to the economy of the world.

It is the increasing value of the information itself that has made this revolution meaningful. For example, an economic value of

information is in its use in decision-making, especially in achieving better use of resources; a social value of information lies in its use to improve the health of the population. So, as I have repeatedly emphasized, the most important issues relate to the effective use of information throughout society.

One concern has been with the relationship between the use of information and the productivity of industry. Does the growth in information activities mean simply increasing "overhead", a drain on the economy, a dissipation of resources in non-productive work? Or does it represent a valuable, even necessary component of production, a positive contribution to the economy, a thing of value both in itself and in the better use of other resources?

Recent research has shown the extent to which productivity is directly related to investment in information resources and services. The continuing growth in the importance of information as a productive component of society has led to increased interest in the burgeoning "information industry". Of special importance in this respect is the report of the Public Sector/Private Sector Task Force of the National Commission on Libraries and Information Science, which addressed several of the issues involved in interaction between the government, the not-for-profit sector, and private enterprise. It clearly supported the view that information resources, products and services were essential to further national development. The Report on the White House Conference on Libraries and Information Services identified the full range of values of information in society, with special emphasis on its importance to the individual.

Most recently, the effect of the new information technologies on

property rights, as represented currently by patents and copyrights, have been the focus of extensive legal debate and wide-ranging discussion. I refer, of course, to the "Betamax Case" which the Supreme Court has been considering since at least January of this year (and it appears will continue to do so for months to come). But beyond that very public litigation, there is deep discussion underway in the publishing community and between publishers and librarians on the implications of the "optical disk" technology on intellectual property rights. As more and more full-text is stored in machine-processible forms, as increasingly data bases are available online or are published as computer-processible files, the problems of balancing the rights of the author, publisher and user are becoming more and more complex. Another issue of current concern in this respect is the relationship between publishers and authors as we move more and more into electronic preparation of manuscripts. What standards should be developed to facilitate that move? What protection does the author have in making an even greater investment personally in the process of production and distribution? What will be the relationship to alternative means for publication and distribution, especially in machine processible forms such as optical disks?

There is even a growing awareness on the international scene of the importance of information economies to national interests. Several countries have adopted "information policies" designed to encourage the development of their own information industries. Although many of the issues relate to computer and communication technologies, the most critical problems relate to information as an

intangible resource, as a part of cultural identity and national pride.

The Educational Implications

So what does that all mean for education? Aside from the obvious need to provide students with "computer literacy" or, perhaps, "information literacy", it seems to me that there is a vital reponsibility that we all have as educators, at every level of education. It's to provide students with the means to evaluate the worth of information to themselves and to society. With such a high percentage of resources in our society being committed to these kinds of products, services and activities, it is absolutely essential that we prepare students today to make the decisions tomorrow, with conscious awareness of what is happening.

The worth of information is difficult to measure. Entrepreneurs involved in developing information resources, products and services have an emotional as well as a financial committment to the belief that they are providing marketable commodities, from which their customers will derive benefits. Librarians and information scientists, involved in the day-to-day operation of information services, have a professional committment to their belief that information is an essential tool. The purveyors of the information technologies--the microcomputers, the optical disks and video recorders, the new communications systems--maintain a steady barrage of advertising hype designed to entice the market to buy their wares, which they describe as the essential means to enter the information age. But the values are by no means clearly identifiable, nor are

they easy to measure.

Personal Values of Information

Certainly, for the individual person, information has immediate value for which we are willing to pay substantial sums of money. We spend years in getting an education that provides us with both a store of information and the tools we need for finding information when we may later need it. We buy books, phonograph records and audio cassettes, cameras and photographs, television sets and video cassette recorders, all for the purpose of getting information.

Some of these personal values are obviouly recreational, since the mind, like the body, needs to have both exercise and relaxation. We read books and view films and watch television because we enjoy the information we receive from them.

And some of the values of information for us as persons lie in the better decisions we can make about our own lives. We get information, in the form of advertising, that helps us decide what products and services we want or need. We get information on plane schedules, hotel reservations, travel opportunities to help us in planning trips for both pleasure and business. We get information about the stock market and real estate so we can invest our money more effectively. We get information about the services of companies and government agencies so we can better use them.

Societal Values of Information

But information is of value to more than just individuals. It provides the primary means for maintaining a society as a cultural entity. Information provides us with our cultural history, our sense of mutual identity, our common purpose.

Information also serves as the means for keeping the several parts of society working together. Communication is obviously necessary for businesses to work together. It's necessary for plane schedules to be maintained. It's part of assuring that goods, materials, food and drink can be distributed throughout our society.

Information is essential to an effective government, especially in a democratic society. Thomas Jefferson said, "The basis of our government being the opinion of the people, the very first objective should be to keep that right; and were it left to me to decide whether we should have a government without newspapers, or newspapers without a government, I should not hesitate to prefer the latter." He also said, "I have sworn upon the altar of God eternal hostility against every form of tyranny over the mind of man." He said so clearly how vital information was to this country!

Economic Value of Information

Information also has economic value, both as a commodity in itself and as a means for better use of other resources. Companies as well as individuals are therefore willing to pay for it.

In their effort to establish the value of information, entrepreneurs and professionals have turned to anecdotal evidence--the experience of users who have benefited from having information readily and reliably available. But the problem is that such evidence is uncertain at best, likely to be apocryphal, and not quantifiable. Rarely does it provide the basis for evaluating the overall effects of the availability of information.

Yet the facts appear to be that our society is steadily increasing the percentage of manpower and other resources being spent

on information. Such growth would seem to substantiate the view that information has value and that the forces of the marketplace have measured that value in their own terms.

In contrast to this, it is also true that investment in information has been an "overhead" expense, and bureaucracies built around the distribution of information represent a drain on the "productive" aspects of the economy. In one analysis, it is shown that over the past 50 years, the costs for information have become a dramatically increasing proportion of the total costs incurred in production of goods.

Thus we have a balancing of views: information as an economic tool, on the one hand, and as an economic drain, on the other. So the problem consists of measuring the balance, in identifying the value of information as a support to the production of other goods and services.

The Qualitative Evidence

There are at least the following six types of qualitative evidence that support the view that information has value as an economic tool:

1) Better work force. It can be argued that information results in a better work force--better trained, more capable of making decisions and dealing with problems, able to adapt better and faster to changing situations. Of course, there may be types of work in which these qualities have no value or even negative value, but those tend to be types of work that become mechanized anyway.

2) Better product development. It can be argued that information results in better product development, since there is more understanding of the needs of the consumer based on information about those needs.

3) Better engineering. Most of the anecdotal evidence on the value of information has focused on the effects of its availability on scientific and technical development, with examples given in which experiments could have been avoided if prior results had been obtained.

4) Better marketing. The most general definition of the "information economy" includes services such as advertising and marketing in general. It is clear that such services lead to improved sales and thus to overall better performance for a company. Beyond that, though, it is also important to note that information about the marketplace is necessary to determine marketing approaches and decisions about allocation of resources. Much of the investment by industry today is in marketing--in information about markets and in advertising to make products known to the markets. The values appear to be evident in the market-based economy of the United States.

5) Better economic data. Decisions made by individual

companies concerning their allocations of resources must be based on information about the economic context within which the companies function. The better the information, presumably, the better the allocation decisions.

6) Better internal management. Finally, it is clear that modern information systems, including the use of both telecommunication and the computer, are valuable and even necessary to good internal management. Porat refers to the "secondary information sector" as the internal information systems of companies and governmental organizations. These are the public and private "bureaucracies" that, while they represent a drain of resources, provide the information that is essential to the management of large organizations.

Balancing these aspects are the following:

1) Evident costs. Most information activities involve very evident costs, in manpower, in equipment, and in purchases.

2) Uncertain return. Rarely are the positive results of any of the "benefits" listed above attributable that clearly to the availability of the information of which they were based. In many cases, the

decisions could have been made without the information; in some cases, they may even be made counter to the information.

3) Long-term return. Even when the value is evident, it is likely that the return is only over the long term, while the expenditure is made in the immediate term. The result is that most information investment must be amortized over a long period of time.

4) Not directly productive. Furthermore, only in rare situations (and most of those in the information industries themselves) is information directly productive. Its value lies in the better uses of other resources, not in the direct contribution to production. Although increasing use of computer-based technologies is changing this situation and increasing the direct contribution to production attributable to information, in the form of programs and data, for most purposes today the role of information is supportive at most.

5) Overhead expense. As a result, in virtually every accounting practice, information is treated as an "overhead" expense, and is therefore subject to all of the cost-cutting attitudes associated with

overhead expense.

6) Differential use. All of the data available on the use of information suggests that most of the use is made by only a few persons: those who know the value and who know how to use the information. As a result, investments that should have wide use turn out to have very limited use.

The Quantitative Evidence

Statistical analyses can be made of data concerning the investment made by industry in information services and their profits. The results show that for every dollar spent for effective, useful information services, industry will make over $2.50 in additional profit. That's a return of two and a half times the investment! Of course, the investments must be properly made, and that requires that the decisions be made by a management that is knowledgeable about the proper use of information.

There are interesting differences in the extent to which different industries in the United States invest in information services. The "high technology" industries--electronics, computers, chemicals, drugs--are the ones that make the greatest investment in information services. They spend over 6.6% of their gross revenues on information services, compared to an average, over all other industry, of 4.4%. That means that the high technology industries spend half again as much on information as do all other industries, on the average.

It is even more interesting to note that those industries that have faced economic difficulties, such as basic steel and the automotive industry, have spent the least on information services. They have averaged about 2.2% of their gross income spent on information services--one third of that spent by the high technology industries!

Although it is difficult to prove that investment in information will result in greater profits, the facts are that those industries which are at the forefront of American productivity are precisely those that spend the most on information. And those that have been in the greatest profit squeeze are those that have been spending the least on information services.

Librarians, Faculty, Administration

I bring to this discussion a particular point of view, represented by the fact that I am dean of a graduate school of library and information science. I see the library as potentially playing a central role in management of information resources, whatever form they may take. Historically, the library has been the institution in society with the defined responsibility for organizing recorded knowledge, providing access to it, and managing the resources needed for those functions. Today, as we move into the expanding world of information, the library has become the central tool in many organizations for information resource management.

Just to mention three examples:

1) In health care, the medical library has been recognized as a vital tool. Now, as a broader range of information resources--patient records, computer

research data files, audio-visual materials--becomes vital to medical research and patient care, the library has been seen as the focal point for information resource management.

2) In the academic research university, the same phenomenon is occurring, with data bases, online computer files, and media resources supplementing the traditional resources of books, journals and reports. Again, the university library is seen by many as the focal point for information resource management in the university.

3) In industry, the same phenomenon is also occurring, and the special library is seen by many (and not just by the librarians) as the focal point for industrial information management, because it provides the means by which external sources of information can be combined with the internal ones that have heretofore been the focus on corporate data processing systems.

In the educational system, we face the same kind of situation, but with an already-established tradition that the library has broader responsibilities, including printed materials, of course, but also including the audio-visual materials. Now we are adding an entire range of computer-based resources. The online data bases are the most

current example, but very soon we will see distribution of massive amounts of materials in computer-processible form--digital video disks, for example.

This, then, is the basis of a new partnership. The library continues to serve as the manager for the school's information resources. That may or may not include the physical hardware, depending upon the specifics of the institutional environment, but it should definitely include the management of the "software"--the information data base in whatever form it may occur.

The responsibility of the faculty in this partnership is then crucial: to incorporate into the curriculum a conscious recognition, far beyond what all of the evidence suggests is currently provided, of the importance of information to the person, to society, to the working environment.

And the responsibility of the administrators? That is clearly to insure that the resources are made available, that they are effectively used, and that all participants are working together in meeting the objectives of preparing students for their lives in an information-rich world.

Conclusion

In conclusion, I want to emphasize one point that I hope has already been made clear, but I am very pedantic and tend to hammer the obvious. It's that the concern with the new information technologies, while of passing importance today, should not be the primary focus of our attention. The technologies, after all, are merely the tools for accomplishing the objectives in using information effectively. Our concentration should be on the objectives, not on the means. That's

the purpose of an education, as contrasted with mere technician training.

Research Foundations of Future Development in Education

Gerald Bracey

Director, Research, Evaluation and Testing

Virginia State Department of Education

Given that the two days of this conference have been rather independent discussions of two articulatable areas, I would like to try and articulate them and I would like to do it thematically, in terms of technology and testing and how they might assist the solution of instructional problems. First I have to lay out a bit of background.

Bill Coffman, in an excellent presentation that told us all that we ought to go back and read a lot of things that were written in the 1930's (and I believe ~~that),~~ asserted that the teachers determine the curriculum, or define the curriculum. That assertion was assented to by many in the audience and by many other speakers, especially Archie La Pointe, who spoke about the one-to-one relationship between teacher and pupil. I tend to disagree about the importance and independence of teachers in this role. I agree more with Dale Carlson, who, in his presentation today, said we probably have something on the verge of a national curriculum because of the uniformity of the textbooks and, consequently, the uniformity of tests. Dale's comments accord very much with the observations of John Goodlad, who has recently concluded that "in the how and the what of instruction, a school is a school is a school."

Closely related to discussions on both days about teachers and their relation to instruction was discussion during the first day of the importance of clinical as opposed to quantitative, formalized

judgments in the teaching and assessing process. I have some real questions about how much that transpires and how good it is. Again, John Goodlad finds in his studies that 70% of all instructional time is teacher talk. Much of the administration activity time of schools is also teacher-dominated, And if talking, as he says, is a good way of organizing your thoughts, then the teachers are doing most of the learning. Goodlad's co-researcher, Kenneth Sirotnik, in a recent Harvard Educational Review, referred to this as the "consistency, persistency and mediocrity of the typical classroom."

But given all this teacher talk and teacher domination of the classroom environment, I doubt seriously that teachers right now are making very good clinical judgments. For one thing, they're not trained for it; in fact, they're probably trained out of it in their preservice programs. Secondly, I don't think they have much time for it. Unless their intuitions allow them to make good diagnoses on brief, informal observations, such as a clinician noticing a rare but very important verbal or behavioral tic, they have little opportunity for developing clinical judment. I back my conclusion up with the observation noted here that high school teachers, who have to deal with hundreds of "clients" each day, put more reliance on tests than do elementary school teachers. But even at the elementary level I am concerned because I am reminded of the Lou Smith studies that found quite a bit of role differentiation across children, but not much role differentiation across time within children, and that childrens' stereotypes tended to perseverate across grades. Teachers in the teachers' lounge would say, "Oh God, yes, I had that person last year, and you can expect this and this and this", and so forth. Certainly

this is a source of information about children, but I'm not sure that it provides a basis for an adequate, informed clinical judgment about them.

Part of the problem with schooling today is the structure of the classroom; 25 to 30 students are just too many to deal with. I am very pleased that Ted Sizer in part of his report already in the June, 1983 Kappan said that same thing, that the structure of the high school is dysfunctional and we need to do something about it. I would point out that these observations accord very well with earlier conclusions about achievement and affective variables reached by Gene Glass and Mary Lee Smith.

With that as background, I want to go on to emphasize that I do believe that the need for accurate clinical judgment is critical, and that's where I think some of the emerging information technologies--microcomputers, the intelligent video disks, the expert systems of software--might be able to help us. I'm not speaking of something as primitive as item banking. I'm not terribly sanguine about item banking: I know of at least one item-banking program where teachers could call for a test from an item bank for any objective they wanted. The people who monitored the system found that they tended to call for tests on things that their children had already mastered, such as phonics, rather than going on to comprehension. In addition, at the 1982 conference on Large Scale Assessment, we had special sessions on item banking, and one theme that emerged was that item banking sounds like a great idea until you do it.

The kinds of things I expect to help are the kinds of programs that were discussed here earlier today, like the Burton and Brown

DEBUGGY and [illegible] programs that look at the patterns of errors that children make and try to diagnose the specific "bug" in the child's problem-solving process. Unfortunately to date, most of these kinds of investigations involve mathematics, because it's a lot easier to write a program to analyze patterns of responses in arithmetic than it is to write a similar program to analyze problems in reading or writing.

There are other disciplines that I think are amenable to what might be called computer-assisted diagnosis. For example, Gordon Novak at the University of Texas has developed an expert system to analyze what kinds of concepts and what kinds of equations are needed to solve certain kinds of physics problems. He finds that problems appearing in textbooks which appear to call for one or two equations actually call for 10 or 12. But the students are not taught 10 or 12, nor are they given an adequate understanding of the concepts in their textbooks. As a recent study by Caldwell found, high school students emerge from physics courses with fundamental misconceptions about Newton's laws of motions. They are, in many instances, left with only their own intuitions or guesses to solve physics problems. The clear implication is that one of the reasons that students are bad at science is they don't have adequate materials to help them understand what's going on. I would hope that the development of future expert systems will help in the diagnosis, assessment and instruction of students in many areas, not just science and mathematics.

A second area where I expect, I hope, that information technology will help in the integral task of assessment and instruction, is in what has been called componential analysis, the attempt to analyze

cognitive tasks into their component parts. Given my affinity for Gestalt psychology, I worry a little that componential analysis may lead to another, more sophisticated round of Titcheneresque structuralism, but even with that reservation I think it's a step in the right direction.

It was mentioned during the first day that kids often ask teachers, "Tell me what I'm doing wrong." If we can get good componential analysis, and I would guess that this will have to be computer-based or computer-assisted, we have an opportunity to assist teachers in answering that question.

I don't think it's going to be that hard, technologically, to do a lot of creative things in componential analysis. I recently saw a drill and practice program that had two real-time clocks built into it. This program ran on a Commodore VIC-20 8K machine. Well, if you can put two real-time clocks on a Commodore VIC-20, imagine what can be done on an Apple or a Commodore 64 or the machines that are certain to be developed in the next few years. Even now it should be possible to analyze reaction time or do some kind of analysis of how long a student lingers over a problem or even a particular part of a problem.

I am also sanguine about using information technology to look at higher-order thinking skills. Dean Gifford said earlier that these investigations were in a relatively primitive state. I would say a very primitive state right now. We are in danger of making great leaps of faith from a few studies based on expert versus novice approaches to physics problems, analyzing life master chess players versus non-master chess players. We need much more of a data base before we start making generalizations or drawing strong conclusions

about the development of expertise. Still, the possibility is there.

This brings me to one topic that has not yet been covered at this conference, which I think is very important and that is that both the areas of assessment and technology need to be informed by developmental psychology. Only a few people have mentioned this, Sam Messick at AERA this year discussed assessment as a developmental construct, but it's the first time I've seen that done recently. I think we are emerging from a Dark Age of psychology. I think our theories of learning have been seriously hampered by psychology's going off and trying to emulate an already defunct model of physics and science back in the 20's and 30's, and I think it's still being hampered somewhat by the fact that most experiments in learning theory are constrained by the convenience of the experimenter, and by the reward structure of universities, which still tend to count number of publications per year.

I think we need to have a general theory of long-term acquisition of competence and expertise over years, and that will require, in addition to the interviewing that Dean Gifford mentioned, a lot of naturalistic and a lot of carefully constructed observation. Otherwise, I think we're in danger of another Dark Age in psychology.

I think it is no accident that today, the giants in psychology are not Thurstone or Gilford or Clark Hull or Neil Miller, they are Freud and Piaget, and Freud and Piaget practically alone. You may disagree with them, and a lot of people do, but they had an observational data base, not an experimental data base. I think that the rise of ethology to respectability, ethology being defined colloquially by Tinbergen as the art of watching and wondering, will

help the people who are doing observational or interview type research. And, as I noted earlier and as Dean Gifford noted, such research is essential.

In my remaining time, I would like to note three problems that concerned me over the course of the conference. I am concerned that cognitive psychology is much too influenced by analogies to digital computers. Many of the overheads dealing with models of problem solving looked like flow charts and even used the symbols of flow-charting. I think we are currently in danger of imposing a model of the brain as a digital computer onto much of our research rather than considering such a model as one of many models and all such models as working hypotheses or metaphors. I don't have time here for a discussion of lateralization theory or the model of the metaphorical mind, but there are many models out there as worthy of consideration as the digital computer.

Secondly, Dexter Fletcher mentioned the invention of the horseless carriage, and I would just like to point out that while we were inventing and reinventing the horseless carriage, we were also experimenting with a wide variety of engine types: steam engines, electrical engines, internal combustion engines, and we have paid dearly for the monopoly that became established by the internal combustion engine. I think that would be the same thing if we had a similar monopoly of one approach to computer-assisted instruction or assessment, and I've even argued that we are paying dearly by having a monopoly in public schooling where a school is a school is a school.

Finally, again in connection with the invention of the horseless carriage, I would like to remind you that Seymour Papert has despaired of educational research contributing much to an educational horseless

carriage because it presumes the traditional classroom. Without a research program that studies children in a variety of environments, we will end up with a set of data that are very much context-bound.

Conceptions of Teaching: A Changing Image

J. Myron Atkin
Dean, School of Education
Stanford University

The emphasis at the conference has been on improving instruction by systematic attempts to enhance the quality of testing and technology. I want in this brief reaction to the excellent papers that have been presented only to highlight another aspect of educational quality: the teacher. My purpose is to raise some possibilities about the impact of developments in testing and technology on our evolving conception of the work of the teacher, and the effect of the changing image on the choice of teaching as a career by some talented people. I believe that the most important factor bearing on the quality of education is the nature of the teaching force. By "nature" I mean the characteristics of those who choose to teach and what those who teach like to do. Do developments in testing and technology change our concept of teaching in a manner that may be unattractive to some able people?

The picture of teaching portrayed at this conference is one of a highly goal-directed activity with a strong emphasis on instructional management. The desired outcomes of schooling are assumed to be very clear, and the task of the teacher seems to be to reach the goals as efficiently as possible. Recent advances in technology and testing have led to a focus almost entirely on tightly specified outcomes.

There is another conception of teaching, that of a versatile and flexible professional who constantly tunes goals as well as techniques to the changing conditions in the classroom and to the varying and

shifting needs of different children. Many goals are worthwhile. The teacher plays a key and sensitive role in selecting those that seem most salient or achievable depending on circumstances. Such a style tends to be attentive to potentially beneficial effects of teaching that are indirect, sometimes long-term, and occasionally incidental---as well as those that are more carefully prespecified. In the papers at this meeting, there seems to be little value placed on such goals. An assumption underlying developments in testing and teaching, so far, seems to be that whatever is worth teaching is worth teaching directly, and the educational result should be apparent quickly.

Of course, many learning objectives, perhaps most, are highly specifiable and short-term. Many, however, are not. What is crowded out of the curriculum when modes of instruction keyed increasingly to tests and computers are emphasized? Is this change pleasing to those who teach or who are contemplating teaching?

Effective teachers have recognized that certain classes of worthwhile educational goals, for example, those like development of sportsmanship, are probably achieved best when the opportune moment arises. If the moment doesn't arise, teach something else. Many thoughtful educators believe that the same is true for many basic concepts in science, social science, and indeed all areas of the curriculum; not all basic concepts, of course, but a significant number. Furthermore, some teachers, some very good ones, relish the joy of the children's educational voyage as much as success in arriving at the educational destination. They derive satisfaction from making choices about the objectives to be stressed at different

times and for different children.

Phillip Morrison, the physicist, used to talk about "enlightened opportunism" in teaching. He stressed the importance of taking advantage of the unexpected event in the clasroom to teach important concepts in science. Indeed, some of the most important ideas in science---ubiquitous concepts like equilibrium, randomness and symmetry---probably are taught best by pointing them out in different contexts when the opportunity arises, rather than striving for such learnings directly. These opportunities often are the incidental results of other pursuits. The very power of the concepts lies in the unexpected richness of their application.

A somewhat related point: Max Beberman, one of the greatest of mathematics teachers, used to try to delay children's verbalization in his teaching to encourage reflectiveness and intellectual discovery. He felt that sometimes statements of conclusions crystallized thought too quickly. He wanted youngsters to ponder various issues in mathematics at leisure. Teaching styles emphasized at this meeting seem inattentive to such an approach.

The major point I am trying to make is that we may be conveying a picture of teaching, with our attention to tests and technology, that de-emphasizes a type of creativity that is valued by, and that has served in the past to attract, an important group of people to the teaching profession, people sensitive to a range of worthwhile educational outcomes poorly adapted to prespecifiation, testing and computers. We may not want to lose this group, and perhaps we can place our efforts in testing and technology in a context that more highly values the importance of continuing to attract such people to the teaching profession.